GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 05:26:48 ; Search time 2436.75 Seconds
(without alignments)
16488.745 Million cell updates/sec

Title: US-09-786-474-1
Perfect score: 1920
Sequence: 1 atgtgcggccttcttggcat.........gctcctacccggtcgagctt 1920

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

rched: 1797656 seqs, 1046326829 3 residues   3595312

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Accession	N	ID	Length	Description
1	1920	100.0	AB029550	1	AB029550	3825	AB029550 Corynebac
2	1904	99.2	AX122490	1	AX122490	1920	AX122490 Sequence
3	1904	99.1	AX127150	6	AX127150	349980	AX127150 Sequence
4	1902.4	99.1	AX063827	6	AX063827	1962	AX063827 Sequence
5	745.4	38.8	AE007071	6	AE007071	14738	AE007071 Mycobacte
6 c	745.4	38.8	270283	1	MTCY190	34150	270283 Mycobacteri
7 c	716.8	37.3	L78813	1	MSGB1551CS	36548	L78813 Mycobacteri
8 c	716.8	37.3	L78814	1	MSGB1554CS	36548	L78814 Mycobacteri
9 c	466.8	24.3	Z99119	1	BSUB0016	215640	Z99119 Bacillus su
10	466.8	24.3	AF008220	1	AF008220	220060	AF008220 Bacillus
11 c	389.2	20.3	AL591980	1	AL591980	103050	AL591980 Listeria
12 c	386.6	20.1	AL596169	1	AL596169	239050	AL596169 Listeria
13 c	245.2	12.8	AL583920	1	MLEPRTN4	348450	AL583920 Mycobacte
14 c	177	9.2	AF168003	1	AF168003	10772	AF168003 Desulfovi
15	116.6	6.1	AB039379	1	AB039379	5188	AB039379 Streptomy
16 c	116.6	6.1	AR031772	6	AR031772	30001	AR031772 Sequence
17 c	114.4	6.1	I32939	6	I32939	30001	I32939 Sequence 1
18 c	105.4	6.0	AJ007747	1	BBR007747	41642	AJ007747 Bordetell
19	101.8	5.5	AE004636	1	AE004636	13280	AE004636 Pseudomon
20 c	99.2	5.3	AE004766	6	RME603646	303100	AE004766 Rhizobium
21	99.2	5.2	AY028431	1	AY028431	14138	AY028431 Burkholde
22	95.8	5.0	AE006451	1	AE006451	34600	AE006451 Lactococc
23 c	95.6	5.0	AL121855	1	SCF62	11070	AL121855 Streptomy
24 c	88.6	5.0	AJ007747	1	BBR007747	34496	AJ007747 Bordetell
25	86.6	4.6	AP003010	1	AP003010	41642	AP003010 Mesorhizo
26	86.6	4.5	Z93940	1	BSZ93940	340857	Z93940 B.subtilis
27	86.6	4.5	Y09476	1	BSY09476	9319	Y09476 B.subtilis
28	85.8	4.5	Z99109	1	BSUB0006	53333	Z99109 Bacillus su
29	85.8	4.5	AL603643	1	RME603643	210440	AL603643 Rhizobium
30 c	85	4.4	AB052749	1	AB052749	260050	AB052749 Corynebac
31	85	4.4	AF195615	1	AF195615	4112	AF195615 Pseudomon
32	82.4	4.3	U47057	1	VCU47057	10508	U47057 Vibrio chol
33	82.4	4.3	Y07786	1	VCLPSS	13171	Y07786 V.cholerae
34	82.4	4.3	AB012956	6	AB012956	16265	AB012956 Vibrio ch
35	77.4	4.0	AX097444	1	AX097444	46721	AX097444 Sequence
36	77.4	4.0	AE004445	1	AE004445	1872	AE004445 Pseudomon
37	74.6	3.9	AJ248286	6	CNSPAX04	10261	AJ248286 Pyrococcu
38 c	74.6	3.9	AX041920	1	AX041920	288440	AX041920 Sequence
39 c	71.8	3.7	AC104736	1	AC104736	349980	AC104736 Pseudomon
40	71.8	3.7	AF035937	1	AF035937	14734	AF035937 Pseudomon
41	69.4	3.6	AF210249	1	AF210249	16736	AF210249 Streptomy
42	68.6	3.6	AP003007	1	AP003007	77457	AP003007 Mesorhizo
43 c	64	3.3	AP003006	1	AP003006	348411	AP003006 Mesorhizo
44	61	3.2	AP001512	1	AP001512	341887	AP001512 Bacillus

ALIGNMENTS

```
RESULT 1
AB029550
LOCUS       AB029550               3825 bp    DNA     linear   BCT 23-MAY-2000
DEFINITION  Corynebacterium glutamicum genes for LtsA, ORF1, complete cds.
ACCESSION   AB029550
VERSION     AB029550.1  GI:6714539
KEYWORDS    ORF1; LtsA.
SOURCE      Corynebacterium glutamicum (strain:KY9611) DNA.
  ORGANISM  Corynebacterium glutamicum
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Corynebacteriaceae;
            Corynebacterium.
REFERENCE   1  (sites)
  AUTHORS   Hirasawa,T., Wachi,M. and Nagai,K.
  TITLE     A mutation in the corynebacterium glutamicum ltsA gene causes
            susceptibility to lysozyme, temperature-sensitive growth, and
            L-glutamate production
  JOURNAL   J. Bacteriol. 182 (10), 2696-2701 (2000)
  MEDLINE   20245524
REFERENCE   2  (bases 1 to 3825)
  AUTHORS   Wachi,M. and Hirasawa,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-1999) Masaaki Wachi, Tokyo Institute of
```

Technology, Department of Bioengineering; 4259 Nagatsuta Midori-ku,
Yokohama, Kanagawa 226-8501, Japan (E-mail:mwachi@bio.titech.ac.jp,
Tel:81-45-924-5770, Fax:81-45-924-5820)
FEATURES   Location/Qualifiers
     source   1..3825
                  /organism="Corynebacterium glutamicum"
                  /strain="KY9611"
                  /db_xref="taxon:1718"
     gene   815..2737
                  /gene="ltsA"
     CDS   815..2737
                  /gene="ltsA"
                  /codon_start=1
                  /transl_table=11
                  /product="LtsA"
                  /protein_id="BAA89484.1"
                  /db_xref="GI:6714540"
                  /translation="MCGLLGLLTANGNAEAFVPALERALPCMRHRGPDDAGTWHDADA
                  AFGFNRLSIIDIAHSHQPLRWGPADEPDRYAMTFNGEIYNYVELRKELSDIGYTFNTS
                  GDGEPIVVGFHHWGESVVEHLRGMFGIAIWDTKEKSLFLARDQFGIKPLFYATTEHGT
                  VFSSEKKTILEMAEEMNLDLGLDKRTIEHYDLQYVPEPDTLHAQISRLESGCTATVR
                  PGGKLEQKRYFKPQPPVQKVVKGEKQDLFDRIAQVLEDSVEKHMRADVTVGSFLSGGI
                  DSTAIAPLAKRHNPDLLTFTTGFEREGYSEVDVAAESAAAIGAEHIVKIVSPEEYANA
                  IPKIMWYLDDPVADPSLVPLYFVAAEARKHVKVVLSGEGADELFGGYTIYKEPLSLAP
                  FEKIPSPLRKGLGKLSKVLPDGMKGKSLLERGSMTMEERYYGNARSFNFEQMQRVIPW
                  AKRENDHREVTAPIYAQSRNFDPVARMQHLDLFTWNRGDILVKADKINMANSLELRVP
                  FLDKEVFKVAETIPYDLKIANGTTKYALRRALEQTVPPHVLHRKKLGFPVPMRHWLAG
                  DELFGWAQDTIKESGTEDIFNKQAVLDMLNEHRDGVSDHSRRLWTVLSFMVWHGIFVE
                  NRIDPQIEDRSYPVEL"
     gene   complement(2882..3226)
                  /gene="orf1"
     CDS   complement(2882..3226)
                  /gene="orf1"
                  /codon_start=1
                  /transl_table=11
                  /product="ORF1"
                  /protein_id="BAA89485.1"
                  /db_xref="GI:6714541"
                  /translation="MTAPSNTGVILTESAASKAKALIDQEGRDDLSLRIAVQPGGCS
                  GLRYQLYFDDRTLDGDKEDIVGGVRLVVDKMSTPYLLGAQIDFADTIEQQGFTIDNPN
                  AGSSCACGDSFN"
BASE COUNT   881 a   1031 c   993 g   920 t
ORIGIN

Query Match   100.0%; Score 1920; DB 1; Length 3825;
Best Local Similarity 100.0%;
Matches 1920; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 atgtgcgggccttcttctggcatattgactgcaaatggaagcgtgaacgctgaagcattcgttcctgca 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   815 ATGTGCGGGCCTTCTTCTGGCATATTGACTGCAAATGGAAGCGTGAACGCTGAAGCATTCGTTCCTGCA 874

Qy   61 ctcgagcggggccttgccatgcatgcgccaccgtggtcctgacgatgccggcacttggcat 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   875 CTCGAGCGGGGCCTTGCCATGCATGCGCCACCGTGGTCCTGACGATGCCGGCACTTGGCAT 934

Qy   121 gacgccgatgcagcgtttggattcaaccgcctctccatcattgatattgcacactcccac 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   935 GACGCCGATGCAGCGTTTGGATTCAACCGCCTCTCCATCATTGATATTGCACACTCCCAC 994

Qy   181 caaccactcggttggggacctgcggatgaaccgaccgacctacgcaatgactttcaacggt 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   995 CAACCACTGCGTTGGGGACCTGCGGATGAACCGACCGACCTACGCAATGACTTTCAACGGT 1054

Qy   241 gagatcctacaactacgttgagctgcgtaaagagtctcggatttgggatataccttttaat 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1055 GAGATCCTACAACTACGTTGAGCTGCGTAAAGAGTCTCGGATTTGGGATATACCTTTTAAT 1114

Qy   301 acttctgggcgatggcgagccaattgtgtcggtttccaccactggcgtttccggtccgtggtc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1115 ACTTCTGGCGATGGCGAGCCAATTGTGTCGGTTTCCACCACTGGCGTTTCCGGTCCGTGGTC 1174

Qy   361 gagcatctccgcggaatgttcggcattgccatttggatacaaaggagaaaagtcgctttc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1175 GAGCATCTCCGCGGAATGTTCGGCATTGCCATTTGGGATACAAAGGAGAAAAGTCGCTTTC 1234

Qy   421 cttgcgctgatcagttcggcatacaagccactgttctacgcaaccaccaccgagcagtcacc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1235 CTTGCGGCTGATCAGTTCGGCATACAAGCCACTGTTCTACGCAACCACCACCGAGCAGTCACC 1294

Qy   481 gtgttcctccagagaagagaacatctgggacatggccggaggatgaatctagatctg 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1295 GTGTTCTCCTCAGAGAAGAAGACCATCTGGACATGGCCGGAGGATGAATCTAGATCTG 1354

Qy   541 ggccttgataagcgcaccattgagcactacgtgacctgagtacgtacgtgccgagccagat 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1355 GGCCTTGATAAGCGCACCATTGAGCACTACGTGACCTGCAGTACGTACGTGCCGAGCCAGAT 1414

Qy   601 acccttcacgcgcagattccgcttggagtcaggctgcaccgcaacagttcgtcgtccgggc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1415 ACCCTTCACGCGCAGATTCCGCTTGGAGTCAGGCTGCACCGCAACAGTTCGTCGTCCGGGC 1474

Qy   661 ggcaagctggaacagaagcgttacttcaagcctcagttcccagtacagaaggtcgtaaag 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1475 GGCAAGCTGGAACAGAAGCGTTACTTCAAGCCTCAGTTCCCAGTACAGAAGGTCGTAAAG 1534

Qy   721 gtaaggagcaggaccctcctcgatcgcatgccaggtgttggaggatagcgtcgaaaag 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1535 GGTAAGGAGCAGGACCAGCACCTCCTCGATCGCATTGCCAGGTGTTGGAGGATAGCGTCGAAAAG 1594

Qy   781 catatgctgccgacgtaggctcgttccttccgcggcgcattgactcaaccgca 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1595 CATATGCTGCCGCCGACGTGCCCGACGTAGGCTGCGTTCCTTCCGCGGCGCATTGACTCAACCGCA 1654

Qy   841 attgccgcttgcaaagcgccaacacctgactcgctcaccttcaccacggtttcgag 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1655 ATTGCCGCCGCTTGCAAAGCGCCAAACCCTGACCTGCTTCACCTTCACCACCGGTTTCGAG 1714

Qy   901 cgtgaaggctactcggagtcgatggcgctatcgcgtctgcgcgctgcgattggcgctgag 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1715 CGTGAAGGCTACTCGGAGGTCGATGGCGCTATCGCGTCTGCGCGCTGCGATTGGCGCTGAG 1774

Qy   961 cacatcgtgaagattgtctcgccctgaggaatacgccaacgcgattcctaagatcatgtgg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1775 CACATCGTGAAGATTGTCTCGCCCTGAGGAATACGCCAACGCGATTCCTAAGATCATGTGG 1834

Qy   1021 tacttggatgatcctgtagctgacccatcattggtccgctgtacttcgtggcagcggaa 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1835 TACTTGGATGATCCTGTAGCTGACCCATCATTGGTCCCGCTGTACTTCGTGGCAGCGGAA 1894

Qy   1081 gcacgtaagcacgtcaaggttgtgctgtctggcgagggcgcagatgagcgtgttcggtgga 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1895 GCACGTAAGCACGTCAAGGTTGTGCTGTCTGGCGAGGGCGCAGATGAGCGTGTTCGGTGGA 1954

Qy   1141 tacaccatttacaaagacccgctatcgcttgctcatttgagaagatccctcccacta 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1955 TACACCATTTACAAAGACCCGCTATCGCTTGCTCCATTTGAGAAGATCCCTCCCACTA 2014

Qy   1201 cgtaaaggcctggaaagctcagcaaggttctgccagacggcatgaaggcaagtccctt 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2015 CGTAAAGGCCTGGAAAGCTCAGCAAGGTTCTGCCAGACGGCATGAAGGGCAAGTCCCTT 2074

Qy   1261 cttgagcgtggctccatgaccatggaagagcgctactacggcaacgtcgctcgctccttcaat 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2075 CTTGAGCGCTGGCTCCATGACCATGGAAGAGCGCTACTACGGCAACGCGCTGCTCCTTCAAT 2134

Qy   1321 ttcgagcagatgcaacgcgttattccatgggacgcaatgggcaaagtgggaccaccgcgaagtc 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2135 TTCGAGCAGATGCAACGCGTTATTCCATGGGACGCAATGGGCAAAGTGGGACCACCGCGAAGTC 2194

Qy   1381 actgcaccgatctacgcacaatccgacacttgatccgattcagtagcccgatgcaacacctg 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2195 ACTGCACCGACCGATCTACGCACAATCCGCAACAATCTTGATCGATTGATCCGATGCAACAACCTG 2254

Qy   1441 gatctgttcacctggatgcgcggcggcacactcctggtcaaggctgacaagatcaacatggcg 1500

```
Qy   181  caacactggttgggaacctgccggatgaaccgcgctacgcatgactttcaacggt  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  CAACCACTGCCTTGGGGACCTGCCGGATGAACCGCGACCGCTACGCAATGACTTTCAACGGT  240

Qy   241  gagatctacaactacgttgagctgcgtaaagagctctcggattgggatatacctttaat  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  GAGATCTACAACTACGTTGAGCTGCGGAGTTTTCAAGGTTGCAAGACC  300

Qy   301  acttcggcgatggcgagccaattgtgtcgttccaccactgggcgagtccggtc  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  ACTTCTGGCGATGGCGAGCCAATTGTGTCGTTCCGGTTCCACCACTGGGCGCGAGTCCGTGTC  360

Qy   361  gagcagattgttccgcctcacgtttcggcattgggatacaaaggaaagtcgctttc  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  GAGCAGATTGTTCCGCCTCACGTTTCGGCATTGGGATACAAAGGAAAGTCGCTTTTC  420

Qy   421  cttgcgcgtgatcagttcggcgtcaacgcactgttctacgcgaaccacgcgagcatgcacc  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  CTTGCGCGTGATCAGTTCGGCGATCAGTTCGGCATTCAAGCCACTGTTCTACGCAACCACCGCATGGCACC  480

Qy   481  gtgttctctcagagagagaccatcttggagatggccgagtgagatgaatctagatctg  540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  GTGTTCTCCTCAGAGAAGAGAAGACCATCTTGGAGATGGCCGAGATGGCCGAGATGAATCTAGATCTG  540

Qy   541  ggcctgataagcgcaccatgagcactacggcagtacggcagtgccgcagccagat  600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  GGCCTGTCGATAAGCGCCACCATGGACACTACGTGGACTACGTGGCAGTGCCGCCGAGCCAGAT  600

Qy   601  accttcacgcgcagattccgcgttggagtcaggtcaggtgcaccgcaacagtcgtcgggc  660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601  ACCCTTCACGCGCAGATTTCCGCGTTGGAGTCAGGTCAGGTGCACCGCAACAGTTCGTCCGGGC  660

Qy   661  ggcaagctggaacagaagcgttactcaagcctcagttcccagtacagaaggtcgtaaag  720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661  GGCAAGCTGGAACAGAAGCGTTACTTCAAGCCTTCAAGCCTCAGTTCCCAGTACAGAAGGTCGTAAAG  720

Qy   721  ggtaaggagcaggagaccctcttcgatcgcattgccaggtgtggaggatagcgtcgaaaag  780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721  GGTAAGGAGCAGGAGACCCTCTTCGATCGCATCGCAGGTGTTGGAGGATAGCGTCGAAAAG  780

Qy   781  catatcgtgccgacgtagcgtcctttccgggcgcattgactcaaccga  840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781  CATATCGTGCCGACGTACCGTGACCGTAGGCTCGTTCCTTTCCGGGCGGCATTGACTCAACCGCA  840

Qy   841  attgcgccgcttgcaaagcgccacaaccctgacctgctcaccttcaccaccggttcgag  900
          ||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841  ATTGCGGGCGCCTTGCAAAGCGCCACAACCCTGACCTGCTCCACCTTCACCACCGGTTTCGAG  900

Qy   901  cgtgaaggctactcggaggtcgatggcgatctcggcgctcgcgatctggcgctgag  960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901  CGTGAAGGCTACTCGGAGGTCGATGTGCTCGATGTGCTCGGCGATTGGCGCTGGCGCTGAG  960

Qy   961  cacatcgtgaagattctcgctgactgaagaatacgcaacgcgattcctaagatcatgtgg  1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961  CACATCGTGAAGATTGTCTCGCCTGGAGGAATACGCCAACGCGATTACACGCCATTCCTAAGATCATGTGG  1020

Qy  1021  tacttggatgatcctgtagctgaccatcattggtccgcgtgactttcgtgcagcggaa  1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021  TACTTGGATGATCCTGTAGCTGACCATGTGGTCCGTGTGTACTTCGTGCGGGCCAGCCGAA  1080

Qy  1081  gcacgtaagcaacgtcaaggtgtgctgtctggcgaggggcgagatgagcgttcggtgga  1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081  GCACGTAAGCACGTCAAGGTTGTGCTGTCTCGCGGAGGGCGCAGATGAGCGTGTTCGGTGGA  1140

Qy  1141  tacaccattacaaagagcgccgtatcgttgctcctccatttggagaagatccctcccacta  1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141  TACACCATTTACAAGAGCCGCCGTATCGTTTCTCTCCATTTGAGAAGAATCCCTTCCCCACTA  1200

Qy  1201  cgtaaggcctggaaagctcagcaaggttctgccagacgcatgaagggcaagtccctt  1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201  CGTAAAGGGCCTGGAAAGCTCAGCAAGGTTCTGCCAGACGGCATGAAGGGCAAGTCCCTT  1260
```

```
Db   2255  GATCTGTTCACCTGCGGATCGCGGCGACACATCCTGGATAAGGCTGACCTGCTCCAAGGCTGCAGAGATCAACATGGCG  2314

Qy   1501  aactccctgagctcgggagttccattcctgttttgGATtcttcaaggttgcagagacc  1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2315  AACTCCCTGAGCTCGGAGTTCCGGAGTTCCATTCCTGGATTCTTCAAGGTTGCAGAGACC  2374

Qy   1561  attccttacgatctgaagattgccaacggtaccaccaagtacgcgcgctgcgcaggcgcactc  1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2375  ATTCCTTACGATCTGAAGATTGCCAACGGTACCACCAAGTACGCGCTGCGCAGGCGCACTC  2434

Qy   1621  gagcagattgttccgcctcacgtttcgaccgtttgcaccgtttggcttcctgttcccatg  1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2435  GAGCAGATTGTTCCGCCTCACGTTTCGACCGTTTGCACCGCTGGGCCGCTTCCCGTTCCCATG  2494

Qy   1681  cgcactggctgccggcgatgagcgtgttcggttgggcgcaggacaccattaaggaatcc  1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2495  CGGCCACTGGCTGCCGGCGATGAGCGTGTTCGGTTCGGTTCGGTTGGGCGCAGGACACCATTAAGGAATCC  2554

Qy   1741  ggtactgaagatatcttcaacaagcaggctgtgctggatatgctgaacgagcaccgcgat  1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2555  GGTACTGAAGATATCTTCAACAAGCAGGCTGTGCTGGATATGCTGGATATGCTGAATGCTGAACGACCGCGAT  2614

Qy   1801  ggcgtgtcagatcattcccgtcgactgtggactgttcgtcatttatgtgtgtgcacggc  1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2615  GGCCTGTCGATCATTCCCGTCGACTGTGGACGTTCTGCTGTTTCGTCATTTATGGTGTGGCACGGC  2674

Qy   1861  atttgtggaaaaccgcattgatccacagatgagaccgctctcacccggtcgagctt  1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2675  ATTTTTGTGGAAAAACCGCCATTGATCAACAGATTGAGGACCCGCTTCCTCACCCGGTCGAGCTT  2734
```

RESULT 2

LOCUS   AX122490   1920 bp   DNA   linear   PAT 11-MAY-2001
DEFINITION   Sequence 2406 from Patent EP1108790.
ACCESSION   AX122490
VERSION   AX122490.1 GI:14039816
KEYWORDS   .
SOURCE   Corynebacterium glutamicum.
ORGANISM   Corynebacterium glutamicum
             Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
             Actinomycetales; Corynebacterineae; Corynebacteriaceae;
             Corynebacterium.
REFERENCE   1 (bases 1 to 1920)
AUTHORS   Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,
             Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.
TITLE   Novel polynucleotides
JOURNAL   Patent: EP 1108790-A 2406 20-JUN-2001;
             KYOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES   Location/Qualifiers
   source   1..1920
             /organism="Corynebacterium glutamicum"
             /db_xref="taxon:1718"
BASE COUNT   422 a   534 c   528 g   436 t
ORIGIN

Query Match   99.2%; Score 1904; DB 6; Length 1920;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1910; Conservative   0; Mismatches 10; Indels 0; Gaps 0;

```
Qy     1  atgtgcggccttctggacatattgactgactgactgacaaatggaagcagcattcgttcctgca  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  ATGTGCGGCCGCCTTCTTGGACATATTGACTATATGACTCAAATGGGAACGCGCTGAAGCATTCGTTCCTGCA  60

Qy    61  ctcgagcgggccttgccatgcatgcgcacgtggtcctgacgatgccggcacttggcat  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  CTCGAGCGGGCCTTGCCGGCCTTGCCATGCATGCGCGCACCGGTGTCTGTCCTCGACGATGCCGGCCACTTGGCAT  120

Qy   121  gacgccgatgcagcgtttggattcaaccgctcgcttcaacgcgcctctccaatcattgatattgcacactcccac  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  GACGCCGATGCAGCGTTTGGATTCGCAGCGTTTGGATTCAACCGCTCGCTTCAACGCGCCTCTCCAATCATTGATATTGCACACTCCCAC  180
```

```
Query Match            99.2%;  Score 1904;  DB 6;  Length 349980;
Best Local Similarity  99.5%;  Pred. No. 0;
Matches 1910;  Conservative  0;  Mismatches  0;  Indels 10;  Gaps  0;

Qy      1 atgtgcggcctcttgcatattgactgcaaatggcaacgtgaagcattcgttcctgca 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228516 ATGTGCGGCCTCTTGCATATTGACTGCAAATGGCAACGTGAAGCATTCGTTCCTGCA 228575

Qy     61 ctcgagcgggccttgccacgtccatgccacccgctagccgcatggcacttggcat 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228576 CTCGAGCGGGCCTTGCCACGTCCATGCCACCCGCTGGTCCTGACCATGCCACACTGGCAT 228635

Qy    121 gacgcgatgcagcgtttggattcaaccgcctccatcatgtattgcacactcccac 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228636 GACGCCGATGCAGCAGCGTTTGGATTCAACCGCCTCCATCATTGATATTGCACACTCCCAC 228695

Qy    181 caaccactgcgtggggacctcggagtgaacccgacctacgcaatgactttcaacgt 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228696 CAACCACTGCGTGGGGACCTCGGAGTGCGTAAAGACCCGACCTACGCCAATGACTTTCAACGGT 228755

Qy    241 gagatctacaactacgttgagctgcaaagagctctcggattgggatatacctttaat 300
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db 228756 GAGATCTACAACTACGTTGAGCTGCGTAAAGAGCTCTCGGATTTGGGATATGCCTTTAAT 228815

Qy    301 acttctgcgatggcgaccaattgtcggttcaccactgggcgcagtccgtggtc 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228816 ACTTCTGCGATGGCGACCAATTGTCGGTTCCACCACTGGGCGGGCCGAGTCCGTGGTC 228875

Qy    361 gagcatctccgggaatgttcggcattggaatacaaaggaaaagtcgctttc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228876 GAGCATCTCCGGGAATGTTCGGCATTGCCATTGGGATACAAAGGAAAAGTCCGTTTC 228935

Qy    421 cttgccgtgatcagttcggcatcaagcgagcaccgttctacgcgagcaccgcgat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228936 CTTGCCGTGATCAGTTCGGCATCAAGCCTGTTCTGGCATTAAGCCACTGTTCTACGCAACCACCGCGAT 228995

Qy    481 gtgttctctcagagaagaaccatctggagatggccgagggagatgaatctagatctg 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228996 GTGTTCTCTCAGAGAAGAAGAAGACCATCTGGAGGATGGCCGAGGGAGATGAATCTAGATCTG 229055

Qy    541 ggccttgataagcgcaccattggaacacgtggacctgcagtacgtgcccgagccagat 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 229056 GGCCTTGATAAGCGCCAGATTGATAAGCGCACCATTGGAACACGTGGACTTGCAGTGCCCGAGCCAGAT 229115

Qy    601 acccttcacgcgcagaattcccgttccccggttcccCTTCCCGGAGTCAGGTca 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 229116 ACCCTTCACGCGCCAGATTCCCGCTTCCCGGAGTCAGGCTCGGCAGGCTGCACCGCACCGCCGGGC 229175

Qy    661 ggcaagtggaacagaagcgttacttcaagcctcagttcccagtacagagaggtcgtaaag 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 229176 GGCAAGCTGGAACAGAAGCGTTACTTCAAGCCTCAGTTCCCAGTTCCCAGTACAGAGAGGTCGTAAAG 229235

Qy    721 ggtaaggagcaggaccttctcgatcgattgccatgccaggtgttggaggtagcgtcgaaaag 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 229236 GGTAAGGAGCAGGACCTCTTCGATCGATTGCCATGCCAGGTGTTGGAGGATAGCGTCGAAAAG 229295

Qy    781 catatggctactcggaggtgtcatggctaggctgagttccggtcggtcatgactcaaccgca 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 229296 CATATGGCTGCTCGGAGGTGTGCGACGTGACCGTAGGCTGTCGTTCCTTCGGCGCCATTGACTCAACCGCA 229355

Qy    841 attgcgccgcttgcaaagcgccacaaccctgtcaacttgttcaccttcaccacccggtttcgag 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 229356 ATTGCGCCGCTTGCAAAGCGCCACAAACCCTGACCTGCTGCACCTGCACCTTCACCACCCGGTTTCGAG 229415

Qy    901 cgtgaaggctactcggaggtcgatggctggctgcggagtccggacgttggcgctgag 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 229416 CGTGAAGGCTACTCGGAGGTCGATGGCTGGGCCATTGGCGGAGTCCGGCGCCGTGCGCCGCTGAG 229475

Qy    961 cacatcgtgaagattgtctcgccgaggaatacgccaacgcgattcctaagatcatgtgg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 229476 CACATCGTGAAGATTGTCTCGCCGAGGAATACGCCAACGCGATTCCTAAGATCATGTGG 229535

Qy   1021 tacttggatgatcctgagctgaccatcattggtcccggctgctacttgctgtgcagcggaa 1080
```

```
Qy   1261 cttgagcgtggctccatgaccatggaagagcgctactacggcaacgctcgctccttcaat 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CTTGAGCGTGGCTCCATGACCATGGAAGAGCGCTACTACGGCAACGCTCGCTCCTTCAAT 1320

Qy   1321 ttcgagcagatgatgccaacgcgttattccatggcaatgggaatcgcaaagcgcaaagtc 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TTCGAGCAGATGATGCCAACGCGTTATTCCATGGCAATGGGAATCGCAAAGCGCAAAGTC 1380

Qy   1381 actgacgatcctacgcacaatcccgacacttgatccagtagccgcatgccgcaacacctg 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 ACTGCGCCGATCTACGCACCACAGTCCCGCAACTTGATCCAGTAGCCGCATGCCGCAACACCTG 1440

Qy   1441 gatctgttcacctggatgcgcggcgacatcctggtcaaggctgacaagatcaacatggcg 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GATCTGTTCACCTGGATGCGCGGCGACATCCTGGTCAAGGCTGACAAGATCAACATGGGCG 1500

Qy   1501 aactccctgagctgcgagttccattcttggataaggaagtttcaaggtgcagagacc 1560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 AACTCCCTGAGCTGCGAGTTCCATTCTTGGATAAGGAAGTTTCAAGGTTCAGAGACC 1560

Qy   1561 attccttacgatctgaagattgccaacggtaccaccaggtacgcgctgcgagggcactc 1620
          |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db   1561 ATTCCTTACGACCTGAAGATTGCCAACGGTACCACCAGGTACGCGCTGCGCCGCCACTC 1620

Qy   1621 gagcagattgttccggcctcacgtttgcaccgcaagagctggcttcccatgtttccatg 1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GAGCAGATTGTTCCGGCCTCCTCACGTTTGCACCGCAAGAGCTGGGCTTCCCGTTCCCATG 1680

Qy   1681 cgccactggcttgccggcgatgagctgttcggttgggcgcaggagacaccattaaggaatcc 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CGCCACTGGCTTGCCGGCGATCGTCTTCGGTTGGGCGCAGGAGACACCATCAAGGAATCC 1740

Qy   1741 ggtgtcagaagtatcttcaacaggcaggctgtgctgactggatatgctgaacgagcaccgcgat 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 GGTACTGCAAGAGTATCTTCAACAGGCAGGCTGTGCTGCTGGACTGTGCTGAACGAGCACCGCGAT 1800

Qy   1801 ggcgtgcagatcattccggctgacttgactgtctgtcatttatggtgtgtggcacggc 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 GGCGTGGTCAGATCATTCCCGTGCGACTGTCGACTGTGTCGTCATTTATGGTGTGTGGCACGGC 1860

Qy   1861 attttgtggaaaaccgcattgatccaacagattgaggaccgctccaccccggttcgagctt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 ATTTTGTGGAAAAACCGCATTGATCCACAGATTGATCCACAGATTGAGGACCGCTCCACCCAGTCGAGCTT 1920
```

```
AUTHORS   Pompejus,M., Kroeger,B., Schroeder,H., Zelder,O. and Haberhauer,G.
TITLE     corynebacterium glutamicum genes encoding metabolic pathway
          proteins
JOURNAL   Patent: WO 0100843-A 109 04-JAN-2001;
          BASF AKTIENGESELLSCHAFT (DE)
FEATURES          Location/Qualifiers
     source        1..1962
                   /organism="Corynebacterium glutamicum"
                   /db_xref="taxon:1718"
     CDS           101..1942
                   /codon_start=1
                   /transl_table=11
                   /note="RXA02139"
                   /protein_id="CAC25155.1"
                   /db_xref="GI:12541540"
                   /translation="MRHRGPDDAGTWHDADAAFGFNRLSIIDIAHSHQPLRWGPADEP
                   DRYAWTFNGEIYNYVELRKELSDLGYAFNTSGDGEPIVVGFHHWGESVVEHLRGMFGI
                   AIWDFKEKSLFLARDQFGIKPLFYATTEHGTVFSSEKKTLEMAEEMNLDLGLGKRFI
                   EHYVDLQYVPEPDTLHAQISRLESGCTATVRPGGKLEQKRYFKPQPPVQKVVKGKEQD
                   LFDRIAQVLEDSVEKHMRADVTVGSFLFGGIDSTAIAALAKRHNPDLLTFTTGFEREG
                   YSEVDVAAESAAIGAEHIVKIVSPPEYANAIPKIMWLDDPVADPSLVPLYFVAAEA
                   RKHVKVVLSGGADELFGGYTIYKEPLSLAPFEKIPSPLRKGLGKLSKVLPDGMKGKS
                   LLERGSMTMEERYYGNARSFNFEQMORVIPWAKREWDHREVTAPIYAQSRNFDPVARM
                   QHLDLFTWMRGDILVKADKIMANSLELRVPFLDEVFKVAETIPYDLKIANGTTKYA
                   LRRALEQIVPPHVLHRKKLGPYPVPMRHWLAGDELFGWAQDTIKESGDIFNKQAVLD
                   MLNEHRDGVSDHSRRLWTVLSFMVHGIFVENRIDPQIEDRSYPVEL."
BASE COUNT      435 a      544 c      534 g      449 t
ORIGIN

Query Match        99.1%;
Best Local Similarity  99.4%;    Score 1902.4;   DB 6;   Length 1962;
Matches 1909; Conservative   0; Mismatches   0; Indels  11; Gaps   0;   Pred. No. 0;

Qy     1 atgtgcggcctcttggcatattgactgacatgggaacgctgaacgcattcgttcctgca   60
         |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db    20 ATGTGCGGCCTCTTGGCATATTGACTGACATGGGAACGCTGAAGCATTCGTTCCTGCA   79

Qy    61 ctcgagcgggcggccttgccatgcatgcgccaccgtggtcctgacgatgccggcacttggcat   120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    80 CTCGAGCGGGCGGCCTTGCCATGCATGCGCCATGCCGTGGTCCTGACGATGCCGACACTTGGCAT   139

Qy   121 gacgccgatgcagcgtttgattcaacgccctccatcattgatattgcacactccac   180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   140 GACGCCGATGGCAGCGTTTGGATTCAACGCCGTTCAACCGCCTCCATCATTGATATTGCACACTCCAC   199

Qy   181 caaccactcgttggggacctgcggatgaaccgaccgtacgcaatgactttcaacggt   240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   200 CAACCACTCGTTGGGGACCTGCGGATGAACCCGACCGTACGCAATGACTTTCAACGGT   259

Qy   241 gagatcctacaactacgttgagctgctaaagagctctcggattgggatatacctttaat   300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   260 GAGATCTACAACTACGTTGAGCTGCTAAAGAGCTCTCGGATTTGGGATATGCCTTTAAT   319

Qy   301 actttggcgatggcggagcaattgtgtcgttccaccactgggcgagtccggtgtc   360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   320 ACTTCTGGCGATGGCGGAGCAATTGTGTCGTTCCGTTCCACCACTGGGCGGAGTCCGGTGTC   379

Qy   361 gagcatctccgcggaatgttcggcattgccattggaatacaaagaaaagtcgctttc   420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   380 GAGCATCTCCGCGGAATGTTCGGCATTGCCATTGGATACAAAGAAAAGTACGCTTTTC   439

Qy   421 cttgcgcgtgatcagttcggcatcaagccgatggttctacgcaaccaccgagcatggcacc   480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   440 CTTGCGCGTGATCAGTTCGGCATCAAGCCGATTATAAGCCACCACTGTTCTACGCAACCACCGAGCATGGCACC   499

Qy   481 gtgttcctcagagaagagaagaccatcttggagatggccgaggagatgaatcctagatctg   540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   500 GTGTTCTCCTCAGAGAAGAAGACCATCTTGGAGATGGCCGAGGAGATGAATGGCAT...GGAGATGAATCTAGATCTG   559

Qy   541 ggccttgataagcgcaccattgagcactgctggacctggacttgcagtcgtacgtgcccccagcccccagat   600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   560 GGCCTTGATAAGCGCACCATTGCACCATTAGCGACTGGACTTGCAGTCGTACGTGGACTTACGCGGCCCCCAGCCCCAGAT   619
```

```
Db 229536 TACTTGGATGATCCTGTAGCTGACCCATCATTGGTCCCGCTGACTTCGTGCCAGCGGAA 229595
Qy   1081 gcacgtaagcacgtcaaggttgtctgctgcagggggcagatgagctgttcggtgga 1140
Db 229596 GCACGTAAGCACGTCAAGGTTGTGTCTGCTGGCGAGGGGCAGATGAGCTGTTCGGTGGA 229655
Qy   1141 tacaccattacaaagaccgctatcgcttgctccattgagaagatccctтcccaccta 1200
Db 229656 TACACCATTACAAGGACGAGCCGCTATCGCCGTTGCTCCCATTGAGAAGATCCCTTCCCACTA 229715
Qy   1201 cgtaaaggcctgggaaagctcagcaagttctgccagacggcatgagaggcaagtccctt 1260
Db 229716 CGTAAAGGCCTGGGAAAGCTCAGCAAGGTTCTGCCAGACGGCATGAAGGGCAAGTCCCTT 229775
Qy   1261 cttgagcgtggctccatgaccatggaagagcgctactacggcaacgctcgcctccttcaat 1320
Db 229776 CTTGAGCGTGGCTCCATGACCATGGAAGAGCGCTACTACGGCAACGCTCGCTCCTTCAAT 229835
Qy   1321 ttcgagcagatgcaacgcgttattccatgggcaaagccgaatggacaccgcgaagtc 1380
Db 229836 TTCGAGCAGATGCAACGCGTTATTCCATGGGCAAAGCGCGAATGGACCACCGCGAAGTC 229895
Qy   1381 actacgcgatctacgcacaatcccgcaactttgatccagtagccgcatgcaacacctg 1440
Db 229896 ACTGCGCCGATCTACGCACAGTCCCGCAACTTTGATCCAGTAGCCCGCATGCAACACCTG 229955
Qy   1441 gatctgttcacctggatgcgcggcgacatcctgtcaagctgacagatcaacatggcg 1500
Db 229956 GATCTGTTCACCTGGATGCGCGGCGACATCCTGGTCAAGGCTGACAAGATCAACATGGCC 230015
Qy   1501 aactcccttgagctgcgagttcattcttggataaggaagtttcaaggttgcagagacc 1560
Db 230016 AACTCCCTTGAGCTGCGAGTTCCATTCTTGGATAAGGAAGTTTCAAGGTTGCAGAGACC 230075
Qy   1561 atccttacgatctgaagattgcaacgtaccaccagactcgctgcgcagggcactc 1620
Db 230076 ATCCTTACGACCTGAAGATTGCCAACGGTACCACCAAGTACGCTCTGCGCAGGGCACTC 230135
Qy   1621 gagcagattgttccgcctcacgtttgcaccgcagaaagctgggcttccctgttcccatg 1680
Db 230136 GAGCAGATTGTTCCGCCTCACGTTTTGCACCGCAAGAAGCTGGGCTTCCCTGTTCCCATG 230195
Qy   1681 cgccactggcttgccggcgatgagctgttcggttgggcgcaggacaccattaaggaatcc 1740
Db 230196 CGCCACTGGCTTGCCGGCGATGAGCTGTTCGGTTGGGCCCAGGACACCATCAAGGAATCC 230255
Qy   1741 ggtactgaagatatcttcaacaagcaggctgtgctggataтgctgaacgagagcaccgcgat 1800
Db 230256 GGTACTGAAGATATCTTCAACAAGCAGGCTGTGCTGGATATGCTGAACGAGCACCGCGAT 230315
Qy   1801 ggcgtgtcgatcatcccgctgactgtggacтgttcтtcatttatgtgtgcacgc 1860
Db 230316 GGCGTGTCGATCATCCCGCTGACTGTGGACTGTTCTTCATTTATGTGTGCACGGC 230375
Qy   1861 atttttgtgaaaaccgcattgatccacagattgaggacegcctcacccggtcgagctt 1920
Db 230376 ATTTTTGTGAAAAACCGCATTGATCCACAGATTGAGGACCGCCTCACCCAGTCGAGCTT 230435
```

```
RESULT 4
AX063827

LOCUS       AX063827        1962 bp    DNA     linear      PAT 24-JAN-2001
DEFINITION  Sequence 109 from Patent WO0100843.
ACCESSION   AX063827
VERSION     AX063827.1  GI:12541539
KEYWORDS    .
SOURCE      Corynebacterium glutamicum.
ORGANISM    Corynebacterium glutamicum
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Corynebacteriaceae;
            Corynebacterium.
REFERENCE   1 (bases 1 to 1962)
```

```
Qy   1681  cgccactggctgccgcgcgatgagctgttcggttgggcgcaggacaccattaaggaatcc  1740
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1700  CGCCACTGGCTGCCGCGCGATGAGCTGTTCGGTTGGGCGCAGGACACCATCAAGGAATCC  1759

Qy   1741  ggtactgaagatatctcaacaagcaggctgtgctggatatgctgaacgagcaccgcgat  1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1760  GGTACTGAAGATATCTTCAACAAGCAGGCTGTGTCTGGATATGCTGAACGAGCACCGCGAT  1819

Qy   1801  ggcgtgtcagatcattccgtcgactggtgccgttctgtcattatgtgtggcacggc  1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1820  GGCGTGTCAGATCATTCCCGTCGACTGGTGGACTGTTCTGTCATTTATGGTGTGGCACGGC  1879

Qy   1861  attttgtggaaaacgcattgatccacagattgaggaccgctcctacccgtcgagctt  1920
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1880  ATTTTTGTGGAAAACCGCATTGATCCACAGATTGAGGACCGCTCCTACCCAGTCGAGCTT  1939
```

RESULT 5

```
AE007071
LOCUS       AE007071        14738 bp     DNA     linear   BCT 27-APR-2001
DEFINITION  Mycobacterium tuberculosis CDC1551, section 157 of 280 of the
            complete genome.
ACCESSION   AE007071 AE000516
VERSION     AE007071.1  GI:13881935
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis CDC1551.
  ORGANISM  Mycobacterium tuberculosis CDC1551
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 14738)
  AUTHORS   Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
            Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
            Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M., Khouri,H.,
            Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
            Gill,J., Mikula,A. and Bishai,W.
  TITLE     Whole genome comparison of Mycobacterium tuberculosis clinical and
            laboratory strains
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 14738)
  AUTHORS   Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
            Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
            Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M., Khouri,H.,
            Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
            Gill,J., Mikula,A. and Bishai,W.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1..14738
                     /organism="Mycobacterium tuberculosis CDC1551"
                     /strain="CDC1551"
                     /db_xref="taxon:83331"
                     /note="clinical strain"
     gene            complement(95..373)
                     /gene="MT2246"
     CDS             complement(95..373)
                     /gene="MT2246"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK46532.1"
                     /db_xref="GI:13881936"
                     /translation="MSQAERNDPMARSGPDTSGGTAPGPGAPVVIDCDDCAARGSGCR
                     DCVVSVLIGVPSLSHDERAALEVLADVGLAPRLRLVPVRRQRSGVA"
     gene            480..2393
                     /gene="MT2247"
     CDS             480..2393
                     /gene="MT2247"
                     /note="identified by match to TIGR protein family HMM
                     TIGR00573"
                     /codon_start=1
```

```
Qy    601  acccttcacggcagagattccgctggagtcgcaggctgcaacagttcgtccgggc  660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    620  ACCCTTCACGGCAGCGCAGATTTCCCGGCTTGGAGTCGCAGGCTGCAACAGTTCGTCCGGGC  679

Qy    661  ggcaagctggaacagagacgttacttcaagctccagttcagtacgagaggtcgtaaag  720
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    680  GGCAAGCTGGAACAGAGACGCGTTACTTCAAGCCTCAGTTCCAGTACAGAGAGGTCGTAAAG  739

Qy    721  ggtaaggagcaggagacctcttcgatcgcattgccaagtgttggaggatagcgtcgaaaag  780
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    740  GGTAAGGAGCAGGAGACCTCTTCGATCGCCATTGCCAGGTGTTGGAGGATAGCGTCGAAAAG  799

Qy    781  catatgcgtgccgacgtaggctgcttccttcggcggccattgactcaaccgca  840
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    800  CATATGCGTGCCGACGTAGGCTAGGCTCGTCGTTCCTTTCGGCGGCCATTGACTCAACCGCA  859

Qy    841  attggcgccgcttgcaaagcggccacacccctgacctgctcacctcaccggtttcgag  900
           ||||||||  |||||||||||||||||||||||||||||||||||||||||||||||||
Db    860  ATTGGCGGCGCTTGCAAAGCCGCCACACCCTGACCTGCTCACCTTCACCCACCGGTTTCGAG  919

Qy    901  cgtgaaggctactcggagagtcgatggctgctgcgggagtccgcccgtgcgattggcgctgag  960
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    920  CGTGAAGGCTACTCGGAGGGTCGATGTGGCTGCTGCGGGAGTCCGCGCCGTGCGATTGGCGCTGAG  979

Qy    961  cacatcgtgaagattgtctcgctgaggaatacgccaacgcgatcctaagatccatcatgtgg  1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    980  CACATCGTGAAGATTGTCTCGCCGGAGGAATACGCCAACGCGATTCCTAAGATCATGTGG  1039

Qy   1021  tacttggatgatcctgtagctgaccgaccatcattggtccggctgtacttcgttgcagcggaa  1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1040  TACTTGGATGATCCTGTAGCTGACCTGTGTAGTTGGTCCGTGTACTTCGTTGCCGCAGCGGAA  1099

Qy   1081  gcacgtaagcacgtcaaggttgtgtctggcgagggggcagatgagctgttcggtgga  1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1100  GCACGTAAGCACGTCAAGGTTGTGCGTCGTGCGGAGGGCGCAGATGAGCTGTTCGGTGGA  1159

Qy   1141  tacaccattacaaagagcacgctatccgtgctgctgctcgatccctccctccccaca  1200
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1160  TACACCATTACAAGGAGCACGCTATCCGTTGCTGCTCCATTTGAAGAGATCCCTTCCCCACTA  1219

Qy   1201  cgtaaaggcctggaaagctcagcaagttctgccagacgcatgaagggcaagtccctt  1260
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1220  CGTAAAGGCCTGGGAAAGCTCAGCAAGTTCTGCCAGACGGCATGAAGGGCCAAGTCCCTT  1279

Qy   1261  cttgagcgtggctcatgaccatggaagagcgctactacggcaacgctcgctccttcaat  1320
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1280  CTTGAGCGTGGCTGCGTCCCATGACCATGGAAGAGCGCTACTACGGCAACCGCTCCGTCCTCCAAT  1339

Qy   1321  ttcggacagatgcaacgcgttattccatgggcaaagcgcgaatgggaccaccgcgaagtc  1380
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1340  TTCGGACAGATGCAACGCGTTATTCCATGGGCAAAGCGCAAAGCGGACCACCGCCAAGTC  1399

Qy   1381  actgcaccgatctacgcacacatcccgcaacttgatccgagccgcatgcaacactg  1440
           |||  |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1400  ACTGCGCCGATCTACGCCACGCCACGTCCCCGCAACCTTGATCCGAGCCCCATGCAACACCTG  1459

Qy   1441  gatctgttcacctgatgcgcggcgacatccggtcaagctgacaagatcaacaatggcg  1500
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1460  GATCTGTTCACCTGATGCGCGGCGACATCCGGTCAAGCTGACAAGATCAACAATGGCG  1519

Qy   1501  aactccctgagctcgagttccattcttcggataaggaagtttcaaggttgcagagacc  1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1520  AACTCCCTTGAGCTCGAGTTCCATTCTTCGGATAAGGAAGTTTTCAAGGTTGCAGAGACC  1579

Qy   1561  attccttacgatctgaagattgccaacggtaccaccaagtacgcgctgcgagggcactc  1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1580  ATTCCTTACGACCTGAAGATTGCCAACGGTACCACCAAGTACGCCGCTGCCGCAGGGGCACTC  1639

Qy   1621  gagcagattgttccgctcacgttttgcaccgcagcggcttccctgttcccatg  1680
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1640  GAGCAGATTGTTCCGCTCACGTTTTGCACCGCAGCGGCTTCCCTGTTCCCATG  1699
```

```
                     /transl_table=11
                     /product="DNA polymerase III, epsilon subunit, putative"
                     /protein_id="AAK46533.1"
                     /db_xref="GI:13881937"
                     /translation="MGATGGTQLSFADLAHAQGAAWTPADEMSLRETTFVVVDLETTG
                     GRRTGNDATPPDAITEIGAVKVCGGAVLGEFATLVNPQHSIPPQIVRLTGITTAMVGN
                     APTIDAVLPMFFEFAGDSVLVAHNAGFDIGFLRAAARRCDITWPQPQVLCTMRLARRV
                     LSRDEAPSVRLAALARLFAVASNPTHRALDDARATVDVLHALIERVGNQGVHTYAELR
                     SYLPNVTQAQRCKRVLAETLPHRPGVYLFRGPSGEVLYVGTRADLRRRVSQYFNGTDR
                     RKRMTEMVMLASSIDHVECAHPLEAGVRELRMLSTHAPPYNRRSKFPYRWWVALTDE
                     AFPRLSVIRAPRHDRVVGPFRSRSKAAETAALLARCTGLRTCTTRLTRSARHGCPACPE
                     LEVSACPAARDVTAAQYAEAVLRAAALIGGLDNAALAAAVQQVTELAERRRYESAARL
                     PVRSAGPWAAWATARSAQLAGEQLSRGWQSDLPTEEHPSREQLFGRTGVDCRTGPPQ
                     RDHLATAIEALWHGQRLRALAALPELIAAKPDCPREGGYQLAVIRHGQLAAAGRAPRG
                     VPPMPVVDAIRRGAQAILPTPAPLGGALVEEIALIARWLAEPGVRIVGVSNDAAGLAS
                     PLLPGRQPFSTAG"
gene             complement(2268..3311)
                     /gene="MT2248"
CDS              complement(2268..3311)
                     /gene="MT2248"
                     /note="similar to PID:1743888; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="anthranilate phosphoribosyltransferase"
                     /protein_id="AAK46534.1"
                     /db_xref="GI:13881938"
                     /translation="MPSWPQILGRLTDNDRDLARGQAAWAMDQIMTGNARPAQIAAFAV
                     AMTMKAPTADEVGELAGVMLSHAHPLPADTVPDDAVDVVGTGGDGVNTVNLSTMAAIV
                     VAAGVPVVKHGNRAASSLSGGADTLEALGVRIDLGPDLVARSLAEVGIGFCFAPRFH
                     PSYRHAAVRREIGVFVNLLGPLTNPARPRAGLIGCAFADLAEVMAGVFAARRSSV
                     LVVHGDGLDELTTTTSTIWRVAAGSVDKLTFDPAGFGFARAQLDQLAGGDAQANAA
                     AVRAVLGGARGPVRDAVVLNAAGAIVAHAGLSSRAEWLPAWEEGLRRASAAIDTGAAE
                     QLLARWVRFGRQI"
gene             3538..4149
                     /gene="MT2249"
CDS              3538..4149
                     /gene="MT2249"
                     /note="similar to GB:D16254 PID:457676 SP:P50677;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="cytochrome c oxidase, subunit III"
                     /protein_id="AAK46535.1"
                     /db_xref="GI:13881939"
                     /translation="MTSAVGTSGTAITSRVHSLNRPNMVSVGTIVWLSSELMFFAGLF
                     AFYFSARAQAGGNWPPPTELNLYQAVPVTLVLIASSFTCOMGVFAAERGDIFGLRRW
                     YVITFLMGLFVLGQAYEYRNLMSHGTSIPSSAYGSVFYLATGFHGLHVTGGLIAFIF
                     LLVRTGMSKFTPAQATASIVVSYWHFVDIVWIALFTVIYFIR"
gene             4190..5032
                     /gene="MT2250"
CDS              4190..5032
                     /gene="MT2250"
                     /note="similar to GP:3080504; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="cytochrome c family protein"
                     /protein_id="AAK46536.1"
                     /db_xref="GI:13881940"
                     /translation="MTKLGFTRSGGSKSGRTRRRLRRLSGGVLLLIALTIAGGLAAV
                     VSTGRMPAMRGEAQAPRKDPIFDEAQIAIGAYVQANGGGPTVVRNPDGSIATQSLRG
                     NDLGRGGDLFRLNCASCHNFTGKGGALSSGKYAPDLAPANEQQILTAMLTGPQNMPKF
                     SNRQLSPEAKKDIIAYKVVATEARQPGGYLLGGFGPAPEGMAMWIIGMVAAIGLALWI
                     GARS"
gene             5029..6318
                     /gene="MT2251"
CDS              5029..6318
                     /gene="MT2251"
                     /note="similar to GP:4539218; identified by sequence
                     similarity; putative"
                     /codon_start=1
```
```
                     /transl_table=11
                     /product="iron-sulfur protein, putative"
                     /protein_id="AAK46537.1"
                     /db_xref="GI:13881941"
                     /translation="MSRADDDAVGVPPTCCGRSDEEERRIVPGPNPQDGAKDGAKATA
                     VPREPDEAALAAMSNQELLALGGKLDGVRIAYKEPRWPVEGTKAEKRAERSVAVMLL
                     GGVFGLALLLFLFWPWEFKAADGESDFIYSLTTPLYGLITFGLSLSIAIGAVLYQKR
                     FIPEEISIQERHDGASREIDRKTVVANLTDAFEGSTIRRKLIGLSFGVGMGAFGLGT
                     LVAFAGGLIKNPWKPVPTAEGKKAVLWTSGWTPRYQGETYLARATGTEDGPPFIKM
                     RPEDMDAGGMETVFPWRESDGDGTTVESHKLQEIAMGIRNPVMLIRIKPSDLGRVVK
                     RKGQESFNEGEFFAFTKVCSHLGCPSSLYEQQSYRILCPCHQSQFDALHFAKPIFGPA
                     ARALQLPTITDTDGYLVANGDFVEPVGPAFWERTT"
gene             6315..7964
                     /gene="MT2252"
CDS              6315..7964
                     /gene="MT2252"
                     /note="similar to GP:4539217; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="cytochrome b, putative"
                     /protein_id="AAK46538.1"
                     /db_xref="GI:13881942"
                     /translation="MSPKLSPPNIGEVLARQAEDIDTRYHPSAALRRQLNKVFPTHWS
                     FLLGEEALYSFVVLLTGVYILTLFFDPSMVDVTYNGVYQPLRGVEMSRAYQSALDISF
                     EVRGGLFVRQIHHWAALMPAAAIMVHLARIFFTGAFRRPRETNWVIGSLLLILAWFEG
                     YFGYSLPDDLLSGLGLRAALSSITLGMPVIGTWLHWALFGGDFPGTILIPRLYALHIL
                     LLPGIITALALGLHLALAVWFOQKHTQFPGQRTEHNVVGVRVMPVFAFKSGAFFAAIVGV
                     LGLMGGLLQINPIWNLGPYKPSQVSAGSQPDFYMMWTEGLARIWPPWEFYFWHHTIPA
                     PVWVAVIMGLVFVLLPAYPFLEKRFTGDYAHHNLLQPRDVPVRTAIGAMAIAFYMVL
                     TLAAMNDIIALKFHISLNATTWIGRIGMVILPPFVYFITYRWCIGLQRSDRSVLEHGV
                     ETGIKRLPHGAYIELHQPLGPVDEHGHPIPLQYQGAPLPKRMNKLGSAGSPGSGSFL
                     FADSAAEDAALREAGHAAEQRALAALREHQDSIMGSPDGEH"
gene             complement(8084..8728)
                     /gene="MT2253"
CDS              complement(8084..8728)
                     /gene="MT2253"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK46539.1"
                     /db_xref="GI:13881943"
                     /translation="MVSRYSAYRRGPDVISPDVIDRILVGACAAVWLVFTGVSVAAAV
                     ALMDLGRGFHEMAGNPHTTWVLAVIVVSALVIVGAIPVLLRARRMAEAEPATRPTGA
                     SVRGGRSIGSGHPAKKRAVAESAPVQHADAFEVAAEWSSEAVDRIWLRGTVVLTSAIGI
                     ALIAVAAATYLMAVGHDGPSWISYGLAGVVTAGMPVIEWLYARQLRRVVAPQSS"
gene             complement(8728..9627)
                     /gene="MT2254"
CDS              complement(8728..9627)
                     /gene="MT2254"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK46540.1"
                     /db_xref="GI:13881944"
                     /translation="MSGPNPPGREPEDEPESEPVSDTGDERASGNHLPPVAGGGDKLPS
                     DQTGETDAYSRAYSAPESEHVTGGPYVPADLRLYDDYEESSDLDDELAAPRWPWVV
                     GVAAIIAAVALVVSVSLLVTRPHTSKLATGDTTSSAPPVQODEITTTTKPAPPPPPAPP
                     PTTEIPTATETQTVTVTPPPPPPATTTAPPATTTTAAAPPTTTTTPTGPRQVTYSV
                     CSIITRAPGDIISVTYYDAAGRRRTQHNVYIPWSMTVTPISQSDVGSVEASSLFRVSKLN
                     CSIITSDGTVLSSNNDGPOTSC"
gene             complement(9813..10232)
                     /gene="MT2255"
CDS              complement(9813..10232)
                     /gene="MT2255"
                     /note="similar to GP:4539223; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical transmembrane protein"
                     /protein_id="AAK46541.1"
```

```
Query Match        38.88;  Score 745.4;  DB 1;  Length 14738;
Best Local Similarity 63.5%;  Pred. No. 1.6e-167;
Matches 1196;  Conservative 0;  Mismatches 666;  Indels 21;  Gaps 3;

Qy     58 gcactgcgagcgggccttgccatgcgatgccaccggtcctgacgatgccggcacttgg 117
          |||||  | | ||||||| ||||  ||  ||||  |||||||||||| | |  |||||
Db  11649 GCCATCGCCCCGCATCGCCATTTGATGCGGCACCCGGGGCCCGATGAATCGGGCACCTGG 11708

Qy    118 catgacgccgatgcagc---------gtttggattcaaccgcctctccatcatt 162
          ||| |||||| |||||| ||       ||  ||||||| |||  ||||||||||
Db  11709 CACGGCCGTCGATGGCGCCTCCGAGGGCTCGTGTTCGGGTTCAACCGACTGTCCATCATC 11768

Qy    163 gatattgcacactccaccaacactgcggtgggacctgcagtgaaccccgaccgctac 222
          ||   || | | | || ||||||| |||| ||||  ||||| | | |||||| |
Db  11769 GACAATCGGCACTCGCCATCAGCCGCTGCGGTGGGGCCGCCGGAGGCTCCGGACCCCTAC 11828

Qy    223 gcaatgactttcaacgtgagatctacaaactacggttgacgcggtaaagagct---ctcg 279
          ||   || ||||||| || |     |  ||| |||| | |||  || ||||    ||||
Db  11829 GTGCTGGTGTTCAACGGCAGGAATCTACAACTACTTCGGAGCTGCGTGACGAGCTGGCACC 11888

Qy    280 gatttgggatatacctttaatactctggcgatgcggagccaattgttgtcggtttccac 339
          |   |||| ||| |  |  |  | ||||| ||||||||||| | |   ||| |||||
Db  11889 CAGCACGGCGCTGTGTTCGCCCACGACGCGTAGGCGACGCGTGACGAGCTCCGGGCTATCAC 11948

Qy    340 cactgggcggagtcggtcggtcgagcatctccggaatgttcggcattgcattgggat 399
          ||||||||  |  ||| |||||||||||||  |||||||  ||||||| ||  |||||
Db  11949 CACTGGGCGCACCGAGTGCTGCAGCCGGTTGCGCCCGCATGTTCGCATTCGCCGGCTGTGGGGAC 12008

Qy    400 acaaaggaaagtcgctttcctggcgtgatcagttcggcatcaagcactgttctac 459
          || |||  |||  | ||     |||| |   ||  || ||||||  ||  | |||
Db  12009 ACCGTCACCCCGCGAATTGTTCTGCGCGAGATCCGTTCGGCATCAAGCCGTTGTTTATC 12068

Qy    460 gcaaccaccgagcgatggcaccgtgttctcctcagaagagaccatcttggagatggcc 519
          || ||||||||||| ||||||||| |||  ||||||| ||| |||| ||||| ||||
Db  12069 GCCACCGGCGGCGGCCCGGTGGCCAGTGAGAAGAAATGCCTGGTCACCCCGTACCACC 12128

Qy    520 gaggagatgaatctagatctgggccttgataagcgcaccattgacactacgtggacctg 579
          ||||||| | ||| ||| | |||||  | ||||| || |||| | |||| |||||||
Db  12129 GAGTTGGTGGGTGGCCGGTTCCGACATCGACCATCGGGCCGTTGCAGCCACTACACCGTCCTG 12188

Qy    580 cagtacgtgcccgagccagataccctgcgcagattccgccgttccagtcaggctgc 639
          ||| |  | ||||  |  | |||  ||||||  | |||| |||| |  |||||||
Db  12189 CAGTACGTGGCCGGAACCCGAGATCGACACCGCACCGGAATCAGGCTGC 12248

Qy    640 accgaacagttcgtccggc---ggcaagctggaacagaaagcgttacttcaagcctcag 696
          ||| || |  | |||  |    || |  |||||| |||  |||  ||||  |||||||
Db  12249 TTCGCCGGATCCGTCCGACCAGCCGCGGGTGATCCGACCGGTGATCACCGGTGATTCGTGCCGCGA 12308

Qy    697 ttcccagtacagaaggtcgtaaaggtaaggagcaggacgaccctcttgatcgatcgccag 756
          |||| | |||||||  | |||||| | ||||  ||||| |  | ||||| ||| | |||
Db  12309 TTTGCGGCCAGTCCGATCACCAACGACAACGACACACGGTGTCACATCGGCCCTATGACAGATCACGGCA 12368

Qy    757 gtgttggaggatagcgtcgaaaagcatatgtgtacttggacgtgccgacgtaggctgttcctt 816
          ||| || || | || |    |    |||  | ||| | | |  ||| |||||| ||||
Db  12369 GTGCGCTTGAGGACTCGGCGTGGCCGCAAGCATAATCGCGGCCGTCACCGTCGGCGCCGGGTTTTCTG 12428

Qy    817 tccggcggcattgactcaaccgcaattggccgttgcaaagcgccacaacccctgaccctg 876
          | |||||||||||| | |||  || ||||||  |||||||  || | |  ||| | |||
Db  12429 TTCGCCGGTATCGACTTCCAGGGTGATCTGCGTGGCCGCCGGCACAATCCGGCGGCTG 12488

Qy    877 ctcacccacacggtttcgagcggttcgactactcggagcgtactccgag 936
          || |||  | | |||  | |||||| |||| | | |||  |  |||||||
Db  12489 ATCACCTTCACCACCGGGTTCCGAGCGGCGAGGGCTTCTCCGAGATCGAGATCGACGTGCGTGGGCT 12548

Qy    937 tccgccgcgattggcggctgagcacatcgtgacctcgtgtctgcctgaggaatagcc 996
          || |||  | |  | ||  ||| | || ||| || ||| | | ||| |||||| ||||
Db  12549 TCGGCAGAGCCGCATCGGGCCGTGCCCAAGGTGTCCACATCGCCCCAAGCCGAGGTCACGAGTTCGTC 12608

Qy    997 aaccgattcctaagatcatgtggtacttggatatcctgtagctgacccatcattggtc 1056
          ||| | | | |    | | || | ||  | |  | || |||  | | |||  | ||||
Db  12609 GCCGCGGCCCTGCGTGGGTTGAGCAGCGCTGACGTGTTGGTACCTTGAGCCGCTGACGAGCTGACCCCAGCCGGTGGTA 12668
```

```
Qy   1057 ccgctgctactcgtggcagcgcgggagcacgtaagcacgtcaaggttgtgctgctggcgag 1116
          |||  ||| | ||| | | ||| ||||| |||| || |||| |||| | || || |||
Db  12669 CCGTTGTTCTTCGCCCGCGGAGGCCGCAAAGCACCATTACAAAGAGACGCCTATCGGTCTGGGGCAA 12728

Qy   1117 ggcgcagtgagctgtcggtggatacaccattacaagacccgtatccgttgctcca 1176
          | ||| ||||||| || |||||||||  ||| |  |  | ||| || ||| |||
Db  12729 GGCGCCGACGAACTGTTCGGCGGCTACACAATCTATCCGAGAACCGCTCGTTGAGGCCG 12788

Qy   1177 tttgagaagatccttcccactacgtaaaggctgggaagctcagcaagttctgcca 1236
          |||  | | | | || | |||| | |  ||||| |||| |  |||  || ||| ||
Db  12789 TTTGACTACCTGCCCAAGCCACTGCCGGCCGGTCGATGGGAAAGTTTCCAAGCCACTGCCG 12848

Qy   1237 gacggcatgaaggcaagtcccttcttgagcgtggtccatgaccatggagagcgctac 1296
          || |||||| ||| |  |||  |||||| | ||| ||| || ||| || |||| ||||
Db  12849 GAGGGCATGGCGCGGCAAGAGTCTGCACCGGGATCGCCGGAGAGCGTCGAAGAGCGCTAC 12908

Qy   1297 tacggcaacgctcgctcccttcaattcgagcagatgcaacgcgttattccatgggcaaag 1356
          || |||||  ||  |  || | |||  || |  ||| ||| | | |  |||  ||||||
Db  12909 TACGGCAATGCCCGCAGTTCTCCGGGCGGCGCAGCTGCGCCCGGTCGAAGTACTGCCCGGGGTTCCGG 12968

Qy   1357 cgcaaatgggaccaccgaagtgagccaatcgtacactacgcacaatccggcaacttgat 1416
          ||  | |||| | |||  | | | |   | |||| | |  |  ||||||||  |  |||
Db  12969 CCGGACTGGACCCACAGCAGTCCCGCCCGTCTACGCGCCGGCCTGGGGAT 13028

Qy   1417 ccagtagcccgcatgcaacacctggatctgttcacctggatcgcggcggcgacatcctgtc 1476
          ||||  | || ||| | | | | ||| | ||||||||| | |  |  || |  || ||| |
Db  13029 CCGGTGGTGGCGGAATGCGAATGCGACACATCGACCTGTTCACCTGGCGCTCGACACGTCTGGTC 13088

Qy   1477 aaggctgacaagatcaacatggcgaacctccctgagctgagctgcggttcccattcttggataag 1536
          | ||| | |||| | |   | || || | |  |  ||| |  |||| ||  |||| ||||
Db  13089 AAGGCCGACAAGATAACGATGGCCAACTCCCTGGAGCTGCGTGGGTGCCGTTCCGTTGGACCCCG 13148

Qy   1537 gaagtttcaaggttgcagagagaccatccttacgatccgaagattgcaacggtaccac 1596
          || || |||   || | |||  ||| | |  | | ||| | ||  ||| || |||||
Db  13149 GAGGTTTTCGCGTGGCCTGGCCTCCCGGTTGCGGCGGCGGCCAAGATCACCCGTACCACACC 13208

Qy   1597 aagtacgcgctgcgcaggcactcgagcagattgttccgcctcacgtttgcaccgcaag 1656
          || | |||| ||| | |||||||| |||  || | |||| ||| | || ||||| |
Db  13209 AAGTACGGCGTGGCGGGCGCGCGCTGGAGCGCTATTGTGTCCGCGACACGTGTCGACCGGCCC 13268

Qy   1657 aagctgggcttccctgttcccatgcgccactggctgccggggatgagctgttcggttgg 1716
          |||| |||| |||  | | ||||| | | | ||||  |||| || | | || ||||||
Db  13269 AAGCTCGGGTTCCCCGATCCGGATCCGGGTACGCCGTGCCGGAGCTGCTGGAGTGG 13328

Qy   1717 gcgcaggacaccattaaggaatccggtactgaagatatctcaacaagcaggctgtgctg 1776
          ||  |||| |  || ||  || || || | | ||  || |  | | ||  | |||  |
Db  13329 GCGTATGCGACGTGGCTCGTCGCGCCAGGCCGGTCGCTCACTCGGTTGCACATCGCCCGGCGTGTAT 13388

Qy   1777 gatatgctgaacgagcagcacgcgcgatggcgtgtcagatcattccgtcgactgtggactgtt 1836
          || | |||||| |  |   |  | |  | || | |  || |  ||| ||| | ||||| |||
Db  13389 CGCATGCTCGACGAGCACACCGGGTTGCGGCGGCAGCAGCCGCCGCGCGCCTGTGTGGACCATG 13448

Qy   1837 ctgtcatttatggtggcacggcattttgtggaaaaccgcattgatcccacagattgag 1896
          ||| |   |  | ||| | || |  | |||  ||  | |||| | || |||| || ||
Db  13449 CTGATCTTTATGCTGTGGCACGCGGATCTTCGTCGCAGCGCAGCAGCGTGGTGCCCGGATCAGC 13508

Qy   1897 gaccgcctcaccggtcgagct 1919
          || |||| |||  ||| |||
Db  13509 GAGCCGCAGTACCCCGTCCAGTT 13531
```

REFERENCE   1 (bases 1 to 34150)
AUTHORS   Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
              Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
              Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
              Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
              Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
              Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
              Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S., and
              Squares,S., Sqares,R., Sulston,J.E., Taylor,K., Whitehead,S. and
              Barrell,B.G.
TITLE   Deciphering the biology of Mycobacterium tuberculosis from the
              complete genome sequence
JOURNAL   Nature. 393 (6685), 537-544 (1998)
MEDLINE   9295987
REFERENCE   2 (bases 1 to 34150)
AUTHORS   Parkhill,J.
TITLE   Direct Submission
JOURNAL   Submitted (11-JUN-1998) Submitted on behalf of the Mycobacter
              tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
              Trust Genome Campus, Hinton, Cambridge CB10 1SA Unite de Genetique
              Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
              75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk

COMMENT   On Jun 27, 1998 this sequence version replaced gi:1237042.
              Notes:
              Details of M. tuberculosis sequencing at the Sanger Centre are
              available on the World Wide Web.
              (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
              been renumbered from the original cosmid submissions but the old
              gene designations are in brackets after the new gene numbers.
              Gene prediction was based on a Hidden Markov Model of TB genes
              implemented in TBparse (Krogh) supplemented with visual inspection
              of positional base preference in codons, especially where there is
              an increase in the observed/expected third position G + C.
              CAUTION: In some cases we may not have predicted the correct
              initiation codon. Where possible we choose an initiation codon
              (atg, gtg, or ttg) which is preceded by an upstream ribosome
              binding site sequence (optimally 5-13bp before the initiation
              codon). If this cannot be identified we choose the most upstream
              initiation codon.

FEATURES   Location/Qualifiers
source   1..34150
                /organism="Mycobacterium tuberculosis H37Rv"
                /strain="H37Rv"
                /db_xref="taxon:83332"
                /clone="Y190"
gene   467..2404
                /gene="Rv2191"
CDS   467..2404
                /gene="Rv2191"
                /note="Rv2191, (MTCY190.02), len: 645, similar to
                SW:DP3A_BACSU P13267 DNA polymerase III, alpha chain
                (31.3% identity in 249 aa overlap) and SW:UVRC_ECOLI
                P07028 excinuclease ABC subunit C (25.7% identity in 230
                aa overlap)"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv2191"
                /protein_id="CAA94260.1"
                /db_xref="GI:1237044"
                /db_xref="SWISS-PROT:Q10384"
                /translation="MQGPNVAAMGATGTGLSFADLAHAQGAAWTPADEMSLRETTFV
                VDLETTGGRTTGNDATPPDAITEIGAVKVCGGAVLGEFATLVNPQHSIPPQIVRLTG
                ITTAMVGNAPTIDAVLMPFEFAGDSVLVAHNAGFDIGFLRAAARRCDITWPQPQVLC
                TMRLARRVLSRDEAPSVRLAALARLFAVASNPTHRALDDARATVDVLHALIERVGNQG
                VHTYAELRSYLPNVTQAQRCKRVLAETLPHRPGVYLFRGPSGEVLYVGTAADLRRRVS
                QYFNGTDRRKRMTEMVMLASSIDHVEGANFRLEAGVRELRMLSTHAPPYNRRSKFPYRW
                WVALTDEAFPRLSVIRAPRHDRVVGPFRSRSKAAETAALLAARCTGLRTCTTRLTRSA
                RHGPACPELEVSACPAARDVTRAQYAEAVLRAAALIGGLDNRALAAAVQQVTELAERR
                RYESAARLRDHLATAIEALWHGQRLRALAALPELIAAKPDGPREGGVQLAVIRHGQLA
                AAGRAPRCVPPMPVVDAIRRGAQAILPTPAPLGGALVEEIALIARWLAEPGVRIVGVS
                NDAAGLASPVRSAGPWAAWAATARSAQLAGEQLSRGWQSDLPTEPHPSREQLFGRTGV
                DCRTGPPQPLLPGRQPPSTAG"
gene   complement(2279..3391)
CDS   complement(2279..3391)
                /gene="trpD"
                /note="Rv2192c, (MTCY190.03c), len: 370. Probable
                anthranilate phosphoribosyltransferase, TRPD, similar to
                eg SW:TRPD_LACCA P17170, (43.2% identity in 308 aa
                overlap), initiation codon uncertain, gtg at 4086
                favoured by homology butthis has no clear ribosome binding
                site"
                /codon_start=1
                /transl_table=11
                /product="trpD"
                /protein_id="CAA94261.1"
                /db_xref="GI:1237045"
                /db_xref="SWISS-PROT:Q10382"
                /translation="MALSAEGSSGGSRGGSGPKAEAASVPSWPQILGRLTDNRDLARQ
                AAWMDQIMTGNARPAQIAAFAVAMTMKAPTADEVGELAGVMLSHAHPLPADTVPDDA
                VDVVGTGGDGVNTVNLSTMAAIVVAAAGVPVVKHGNRAASSLSGGADTLEALGVRIDL
                GPDLVARSLAEVGFCFAPRFHPSYRHAAVVRREIGVPVFNLLGPLTNPARPRAGL
                IGCAFADLAEVMAGVFAARRSSVLVVHGDDGLDELTTTTTSTIWRVAAGSVDKLFDP
                AGFGFARAQLDQLAGGDAQANAAAVRAVLGGARGPVRDAVVLNAAGAIVAHAGLSSRA
                EWLPAWEEGLRRASAAIDTGAAEQLLARWVRFGRQI"
gene   3549..4160
                /gene="ctaE"
CDS   3549..4160
                /gene="ctaE"
                /note="Rv2193, (MTCY190.04), len: 203. Probable ,ctaE,
                strong similarity to cytochrome c oxidase polypeptide III
                (cox3) eg SW:COX3_SYNY3 Q06475 (29.8% identity in 225 aa
                overlap)"
                /codon_start=1
                /transl_table=11
                /product="ctaE"
                /protein_id="CAA94262.1"
                /db_xref="GI:1237046"
                /db_xref="SWISS-PROT:Q10385"
                /translation="MTSAVGTSGTAITSRVHSLNRPNMVSVGTIVWLSSELMFFAGLF
                AFYFSARAQAGGNWPPPPTELNLYQAVPVTLVLIASSFTCQMGVFAAERGDIFGLRRW
                YVITFLMGLFFVLGQAYEYRNLMSHGTSIPSSAYGSVFYLATGFHGLHVTGLIAFIF
                LLVRTGMSKFTPAQATASIVVSYVWHFVDIWVIALFTVIYFIR"
gene   4201..5043
                /gene="qcrC"
CDS   4201..5043
                /gene="qcrC"
                /note="Rv2194, (MTCY190.05), len: 280. Probable cyoA,
                shows similarity to cytochrome c family; contains 2 X
                PS00190 Cytochrome c family heme-binding site signature"
                /codon_start=1
                /transl_table=11
                /product="qcrC"
                /protein_id="CAA94263.1"
                /db_xref="GI:1237047"
                /db_xref="SWISS-PROT:Q10386"
                /translation="MTKLGFTRSGGSKSGRTRRRLRRLSGGVLLLIALTIAGGLAAV
                LTPTPQVAVADESSSALLRTGKQLFDTSCVSCHGANLQGVPDHGPSLIGVGEAAVFQ
                VSTGRMPAMRGEAQAPRKDPIFDEAQIDAICAYVQANGGGPTVVRNPDGSTATQSLRG
                NDLGRGGDLFRLNCASCHNFTCKGGALSSGKYAPDLAPANEQQILTAMLTGPQNMKF
                SNRQLSFEAKKDIIAYVKVATEARQPGGYLLGGFGPAPEGMAMWIIGMVAAIGLALWI
                GARS"
gene   5040..6329
                /gene="qcrA"
CDS   5040..6329
                /gene="qcrA"
                /note="Rv2195, (MTCY190.06), len: 429. Probable cyoB,
                shows some similarity to cytochrome B6-F complex
                iron-sulphur subunits (Rieske iron-sulfur protein);
                contains PS00200 Rieske iron-sulfur protein signature 2"
                /codon_start=1
                /transl_table=11
                /product="qcrA"
                /protein_id="CAA94264.1"
                /db_xref="GI:1237048"
                /db_xref="SWISS-PROT:Q10387"

```
        /translation="MSRADDDAVGVPPTCGGRSDEEERRIVPGNPQDGAKDGAKATA
        VPREPDEAALAAMSNQELLALGKLDGVRIAYKEPRWPVEGTKAEKRAERSVAVWLLL
        GGVFGLALLILFLFEWPWEFKAADGSDFIYSLITTPLYGLTFGLSLSIATAGVLYQKR
        FIPEISIQERHDGASREIDRKTVVANLTDAFEGSTIRRRKLICLSFGVGMGARFGLGT
        LVAFAGGLIKNPWKPVPVPTAEGKAVLWTSGWTPRYQGETIYLARATGTEDGPPFIKM
        RPEDMDAGGMETVFPWRESDGDGTTVESHHKLQEIAMGIRNPVMLIRIKPSDLGRVVK
        RKGQESFNPGEFFAFTKVCSHLGCPSSLVEQQSVRILCPCHQSQFDALHFAKPIFGPA
        ARALAQLPTIIDTDGYLVANGDFVEPVGPAFWERTT"
gene    6326. .7975
        /gene="qcrB"
CDS     6326. .7975
        /gene="qcrB"
        /note="Rv2196, (MTCY190.07), len: 549. Probable cytB,
        integral membrane protein, low similarity in
        amino-terminal half to cytochrome b subunits, highly
        similar at C-terminus toSW:12KD_MYCLE P15878 12 KD protein
        PIR:S08427 (86.9% identity in 153 aa overlap). FASTA
        scores: sp|Q45658|QCRB_BACST MENAQUINOL-CYTOCHROME C
        REDUCTASE (224 aa) Opt: 341 z-score: 402.4 E(): 6.8e-15;
        28.0% identity in 207 aa overlap"
        /codon_start=1
        /transl_table=11
        /product="qcrB"
        /protein_id="CAA94265.1"
        /db_xref="GI:1237049"
        /db_xref="SWISS-PROT:Q10388"
        /translation="MSPKLSPPNIGEVLARQAEDIDTRYHPSAALRRQLNKVFPTHWS
        FLLGEIALYSFVLLITGYLTLFFDPSMVDVTYNGVYQPLRGVEMSRAYQSALDISF
        EVRGLFVRQIHHWAALMFAAAIMVHLARIFFTGAFRRPRETNWVIGSLLLILAMFEG
        YFGYSLPDDLLSGLGLRAALSSITLGMPVIGTWLHWALFGGDFPGTILIPRLYALHIL
        LLPGIILALIGLHLALVWFQKHTQFPGPGRTEHNVVGVRVMPVFAFKSGAFFAAIVGV
        LGLMGGLLQINPIWNLGPYKPSQVSAGSQPDFYMMWTEGLARIWPPWEFYFWHHTIPA
        PVWVAVIMGLVFVLLPAYPFLEKRFTGDYAHHNLLQRPRDVPVRTAIGAMAIAFYMVL
        TLAAMNDIIALKFHISLNATTWIGRIGMVILPPFVYFITYRWCIGLQRSDRSVLEHGV
        ETGIIKRLPHGAYIELHQPLGPVDEHGHPIPLQYQGAPLPKRMNKLGSAGSPGSGSFL
        FADSAAEDRALREAGHAAEQRALAALREHQDSIMGSPDGEH"
gene    complement(8152. .8796)
        /gene="Rv2197c"
CDS     complement(8152. .8796)
        /gene="Rv2197c"
        /note="Rv2197c, (MTCY190.08c), len: 214. Unknown"
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein Rv2197c"

Query Match    38.8%;    Score 745.4; DB 1;    Length 34150;
Best Local Similarity   63.5%;   Pred. No. 1.7e-167;
Matches 1196;   Conservative   0;   Mismatches 666;   Indels   21;   Gaps   3;
```

```
Qy     58 gcactcgagcgggcttgccatgcagccaccgtggtcctgacgatgccggcacttgg  117
          ||  |||| |||||||||||||||||  || ||||||||||| ||||||| |||
Qy  11717 GCCATCGCCCGCGAATCGCATTGATGCGGCACCGCCCGATGAATCGGGACCACCTGG 11776

Qy    118 catgacgacgatgcagc----------gttttgattcaaccgcctctccatcatt  162
          ||  ||||||| ||||||           ||||||||| |||||||||||||||
Db  11777 CACGGCCCGTCGATGGGCCGCCAGGCGTCGGGTTCAACCGCCGCTGTTCCATCATC 11836

Qy    163 gatattgcacactcccacaccactgcttgggaggaccctgcgggatgaaccgcgctac  222
          ||  |||||||||||||| |||| |||||||||||||| |||||  |||||||||||
Db  11837 GACATCGCGCACACGCGCGACGGCGACTGCTCGAGGCGGCGGCCGTGCGCGGTGCTAC 11896

Qy    223 gcaatgactttcaacggtagatctacaactaactacgttgagagct----ctcg  279
          ||  |||| |||| |||||||| ||||  ||| |||||| |||||    |||
Db  11897 TACGGCACCTGTTCAACGGCAGATCGCACAGTCGCGCGCGGAAGTACTGCCCGGTTCCGG 11956

Qy    280 gatttggatataccttaatacttctggcgatggcgagccaattgttgtcggtttccac  339
          ||  ||||| ||||| |||||||||  |||||| |||||| ||||  ||| |||||||
Db  11957 CAGACGCGGCTGTGTTCGCCACCACGGCGGCGATCCTCACCGCGATCTATCAC 12016

Qy    340 cactggggcgagtccgtcgtggacatccccgcgggaatgttcggcattgccattgggat  399
          ||  ||||| |||||||||||||| ||||||||  |||| ||||||||| |||||  ||
Db  12017 CACTCGGGGCACCGAGGTGCTGCAGGCGTGCGCCGGTTCGCATTCGGCGTGTGGGAC 12076
```

```
Qy    400 acaaaggaaagtcgctttcctggcgtgatcagttcggcgatcaagccactgttctac  459
          ||  |||  |||||||  ||  ||||||||| ||||| |||||
Db  12077 ACCGTCACCCGCGAATTGTTCTGCGGCGGAGATCCGTTCGGCGACATCAAGCCGTTGTTTATC 12136

Qy    460 gcaaccacgacgatgcaggcaccgtctcctccagagaagagacacctggagatggcc  519
          ||  |||||| |||||||| ||||||||| | |||||  ||  |||||||||||||
Db  12137 GCCACCGGAGCCGGCCGGCGGGTGGCCAGTGAGAAGAAATGCCTGCTGACACCTCGTC 12196

Qy    520 gaggagatgaatctagatctgggccttgataagcgcaccattagcactacgtggacctg  579
          ||  |||||||  |||||| |||||| || ||||| |||||||| || |||||||||
Db  12197 GAGTTGGTGGGGTTCGACACCGGATCGACCATCGGCGTTGCAGCACTACACCGTCCTG 12256

Qy    580 cagtacgtgccgcgagccagcagataccctcacgcgcagatttccgcgtcggagtcaggctgc  639
          ||  |||||||||||||  ||||||| | |||||||||| |||||||||||||||| | ||
Db  12257 CAGTACGTGCCGCGAACCCGGGGTACGTCGGGGTACGTCGGCCTGGAATCAGGCTGC 12316

Qy    640 accgcaacagtctgtccggc---ggcaagctggaacagaaagcgtactccagcctcag  696
          ||  ||||| |||||||||||   ||  ||||| |||| ||| |||||||||  |||
Db  12317 TTCGCCCGGATCCGTGCCGACCAGCTCGCGCCGTGATCACCCGGTTATTTCGTGCCGCGGA 12376

Qy    697 ttcccagtacagaaggtcgtaaagggtaaggacaggacctcctcgatcgcattgcccag  756
          ||  || ||||| || |||| ||||| |||| |||| |||||||| || || |||||||
Db  12377 TTTGCGGCCAGTCCGATCACCACCAACGACCAGGCCCGTTGCAGCGAGATCACGGCA 12436

Qy    757 gtgttggaggatagcgtcgaaaagcatatgcgtgcgacgtgaccgtaggctcgttcctt  816
          ||  || ||||||| ||||||||  |||| |||||||||||||  ||||||  | ||||
Db  12437 GTGCTTGAGGACTGGGTGCCAAGCACCATGCGGCCCGATGTCACCCGGCCGTTTCTG 12496

Qy    817 tccggcggcattgactcaaccgcatcggccgcttgcaaagcgccacaaccctgacctg  876
          ||  ||||||||||| ||| |||| ||||||||||| ||||||||||||| |||||||
Db  12497 TCCGGGGTATCGACTCCGCCAGCTTCGGCCATCGGCCATCCGGCACAATCCGCGGCTG 12556

Qy    877 ctcacccttcaccacggttcgagcgtgagcacatcggaggtgctactcggaggtgctggcgag  936
          ||  |||| ||| || |||||||| ||| || |||||| || |||||| ||||||| || || |
Db  12557 ATCACCTTCACCACCACCGGTTTCGAGCGCGACGATCGGACGTCGCGCGGTGCT 12616

Qy    937 tccgccgctgcgattggcgctgagacatcgtgaagattgtctcgcctgaggaatacgcc  996
          ||  |||| |||| ||||| |||  |  |||| |||  || ||| | |||||||  ||
Db  12617 TCGGCAGAGGCCATCGGGTCGCCGCTCACTCGCCAAGGTGGTCAGCGACGAGTTCGTC 12676

Qy    997 aacgcgattcctaagatcatgtggtacttggatcctgtagctagccaccattggtc 1056
          ||  ||| |||| ||||| ||||| |||| |||||||  |||| || |||||||  |
Db  12677 GCCGCCCTGCCCGAGATCGTCTGTACCTCGACGAGCCGGTCGCGACGAGCCGCTGTA 12736

Qy   1057 ccgctgtacttcgtggcgacggaagcacgtaagcacgtcaaggttgtgtctggcgag 1116
          ||  ||| |||| |||||||| |||||||||  |||| |  |||||||| |||||
Db  12737 CCGTTGTTCTTCGTCGTGGCGGACGGCCGAGGCCCGAAAGCCCGAAAGTGGTTGTCGGGGCGAA 12796

Qy   1117 ggcgcagatgagctgttcggtggatacaccattacaagagccgtatccgcttgctcca 1176
          ||  ||| ||||||| |||||||| |||| || |||||||||||  | |||||| ||
Db  12797 GGCGGCCGACGAACTGTTCGGGCGGCTACACAATCTATCGAGAACGCGTGTGTCGTTGAGGCCGG 12856

Qy   1177 tttgagaagatccctcccacactacgtaaaggctgggaaagctcaggaaggttctgcca 1236
          ||  || || |||| ||||| || ||  |||| |||||||||| |||| ||| ||||
Db  12857 TTTGACTACCTGCCCAAGCCACCGCTCGATGGGAAAAGTTCCAAGCCACTGCCG 12916

Qy   1237 gacggcatgaaggcaagtccctcttcttgagcgtggccatgaccatgaagagcgctac 1296
          ||  |||||||||||| ||| | || ||||  ||||||| ||| | ||||||| |||
Db  12917 GAGGGCATGCGCCGGCGCAAGAGTCTGCTCGGATCGGCGATCGCTGCACACTGCCACTTAC 12976

Qy   1297 tacggcaacgctcgctccttcaattcgaagagagatgcaacgcgttattccatgggcaaag 1356
          ||  |||| |||||||| |||||| ||  ||||||||| ||||||| ||||||| |||||
Db  12977 TACGGCAATGCGCGGCCAGCTGTCCGGCGGGAAGTACTGCCCGGTTCCGGACC 13036

Qy   1357 cgcgaatggggatatacctcgaagtgctggcgatggcgagccaccgaccgatcacgcacaat cccgcaactccgc  1416
          ||  ||||| ||||||| ||  |||| | |||  |||| ||||||| ||||||| ||||  ||||||  |||||||
Db  13037 CCGGACTGGACCCCACACAGATGTCACGGCGCGGTCTACGGCGGATCCTCCGCCGGCTATCAC 13096

Qy   1417 ccagtagccccgatcgacaacctggtctgttcacctgatgtcgcggcgacatcctgtc 1476
          ||  ||| |||||||||||||||| | ||||||| |||| ||| |||||| ||| |||
Db  13097 CCGGTGCGGCCGGCACCGGAATGCAGCACATCAGCAGCACCTGTTCACCTGCGGACGACATTCTGTC 13156

Qy   1477 aaggctgacaagatcaacatggcgaactccctgacctgcggtcgcgaagttcattcttggataag 1536
          ||  ||| ||||| |||| |||||||| ||||| | || ||  | |||| |||||||||||||||
```

another
clone. Coding sequences larger than 60 amino acids were predicted
on
the basis of codon usage and homology information. An attempt was
made
to locate the most probable start site based on codon usage,
homology,
the presence of a Shine-Dalgarno sequence, or overlapping orf that
suggested translational coupling. It is possible that the actual
start
site differs from the one selected.

```
FEATURES             Location/Qualifiers
     source          1..36548
                     /organism="Mycobacterium leprae"
                     /specific_host="Dasypus novemcinctus"
                     /db_xref="taxon:1769"
                     /clone="cosmid B1551"
                     /tissue_lib="Lorist 6"
                     /note="The liver of the armadillo was used to isolate the
                     Mycobacterium leprae."

BASE COUNT     6874 a   10335 c   11630 g    7709 t
ORIGIN

Query Match          37.3%;      Score 716.8;   DB 1;   Length 36548;
Best Local Similarity  63.3%;    Pred. No. 1.2e-160;
Matches 1134;   Conservative   0;   Mismatches 652;   Indels   6;   Gaps   2;

Qy    133 gcgtttggattcaaccgcctctccatcattgatatggcacactcccaccaccactgcgt  192
          ||||||||| |||  ||| ||| ||| ||| || || || ||||||   |||||||||||
Db  11406 GCGTTTGGGTTTAACCGGTTGTCCATCATCGACATTCGCATCAGCCGCTGCGG  11347

Qy    193 tggggacctcgcggatgaaccgaccgtacgcaatgactttcaacgtgggatctcacaac  252
          |||||||| |  ||| || ||| || || ||| |||| |||||  ||||||| ||||||
Db  11346 TGGGGACCACCGGAGACACCACCCAACCGGTACAATCCTGGTATTCAACGGCGAGATCTACACAAC  11287

Qy    253 tacgttgagctgcgtaaagagc---tctcggagttggatatacctttaatacttctggc  309
          |||| |||||||||| |  ||||   | ||  ||| ||| |  ||  ||||||  |||||
Db  11286 TACCTGGAGCTGCGGCGACGAGCTGGTCATCCGGCACGCGCTGCTTTCGCCACCGGACGGA  11227

Qy    310 gatgcgagccaatgttgtcggttccaccactggggcgagtccgtggtcgagcatctc  369
          ||  ||||||| || |||| ||| ||||||| ||  | ||   |||  |||||||| |
Db  11226 GACGGTGAATCTATCGTCGCCGGCTTCCATTACTGGGCACCGGAGATACTAACACGGCTG  11167

Qy    370 cgcggaatgttcggcattgcattggagtcacatttggatacaaaggaaaagtcgctttcttcttgcgcgt  429
          |||||  |||||||| || |  || || |||| | || ||  ||||||||||| |||||| ||  |||||
Db  11166 CGTGGCATGTGTTCGCCTTCGCCGTGTGGGACACCATCGCCCACGAATTGTTCTGTGCCCGA  11107

Qy    430 gatcagttcgcatcaagccactgttctacgcaaccacgagcatggcaccgtgttctcc  489
          ||||| ||||| ||||| |||||||||| |  ||||| ||| ||||||||| ||| ||
Db  11106 GACCCTTTCGGGATCAAACCACTGTTCATGGCGACCGGACCCGCACCGCGGTGCC  11047

Qy    490 tcagagaagaagaccatcttggagatggccgagaggagtgaatctagatctgggccttgat  549
          | ||||||| || ||||| |||||| ||||||  ||||| || || || ||||| ||||||
Db  11046 AGCGAAAAGAAAATGCCTGCTTGAACTGGCCGAAGTGATGATCAGATTCGACACGCAGATTGAC  10987

Qy    550 aagcgcaccattgagcactacgtgacctgcgtcagtacgtgcccgagccagatacccttcac  609
          || ||| ||||| ||||||||| |||| |||| |||| || |||||| ||||| ||  |||
Db  10986 ACCCGCGCACTACAGCACTACACGGTCCTGCAGTACGCTGCCCGAGCGCAGCGCTGCAT  10927

Qy    610 gcgcagatttcccgcttggagtcaggctgcaccgcaacagttcgtcaccgcgcg---caag  666
          || |  ||||||||| ||||| || || ||||| | ||||||| |   |||||    |||
Db  10926 CGCGGGGTGCACCGACTCACCGACTCGAATCTGGCTGTGTTACGCACGTATCCGTCCGAACAGCTCGAT  10867

Qy    667 ctggaacagaagcgttacttcaagcctcagttcccagtacagaaggtcgtaaaggtaag  726
          ||||||| |||| | |||||| |  || |||||||| || | ||||   |||||  |||
Db  10866 CCAGCGACCACACCGGCCACCCGCTACTTCCGTGCCCAATTGCCCCACCGATCACCGAGCACT  10807

Qy    727 gagcaggacctcttcgatcgatcgattgcccaggtgttgaggatagcgtcgaaaagcatatg  786
          || |||||||| ||||| ||||| |||||| |||||||||||| ||||| |||||||||||
Db  10806 GAACAGACCCGCTACGGCGAAATCACCGCAGTGCTTGCGGACTCAGTGCTTGGCCAAGCACATG  10747

Qy    787 cgtgccgacgtgaccgtaggcgtgttccttccgcggcgcattgactcaaccgcaattgcg  846
```

```
Db  13157 AGGCCGACAAGATAACGATGGCCAACTCCCTGGAGCTGCGGGTGCCGTTCCTGGACCCG  13216
          |||||  |||||||  ||| ||||||||||||||  |||||| |||  ||| |||||||
Qy   1537 gaagtttcaagttgcagagaccattccttacgatctgaagattgccaacggtaccacc  1596
          || ||||| |  ||  |  ||   ||  |   ||  |||||||| |  |||  ||||||
Db  13217 GAGGTTTCGCGGTGGCCCTCCCGGTGCCGGCGGCGGCCAAGATCACCCGTACCACCACC  13276

Qy   1597 aagtacgcgctcgcgcagggcactcgaacagagattgttccgcctcacgtttgaccgcaag  1656
          ||||||| || | || ||| | ||  | ||||  |||||||| ||||| || |||||||
Db  13277 AAGTACGGCGCTGCGGCGCGGCGGCCTATTGTGCCCGCCACACGTGCTGCACCGGCCC  13336

Qy   1657 aagtggggcttccctgttccatggcgccactggctgctgccgcgatgagtgttcggtgg  1716
          ||||| ||||| ||| ||||| |||||||  |||| || ||||| ||||| ||||||||
Db  13337 AAGCTCGGGTTCCCGGTTCCCGGCCATTCCGGCATCCGGCCTGCTGCGGCAGCTGCTGGAGGTGG  13396

Qy   1717 gcgcaggacaccattaaggaatccggtactgaagatatcttcaacaagcaggcgtgctg  1776
          ||  || ||||||||||| |||| || | || ||||||| || ||| ||||| || ||
Db  13397 GCCTATGCGACGGTGGCCTCGTCGCCAGGCCGGTCACTTGGTTGACATCGCCGCCGTGTAT  13456

Qy   1777 gatatgtgaacgagcaccgcgatggcgtgtcagatcattcccgtcgactggactgtt  1836
          || ||| |||||||||||| || ||||  |||||||| ||||||||||||||| |||
Db  13457 CGCATGCTCGACGAGCACCGGTGCGGCAGCAGCGACCACACCGCCGGCTGTGCGACCATG  13516

Qy   1837 cgtcatttatgtgtgcacggcattttgtgaaaaccgcattgatccacagattgag  1896
          |||||||||||| || ||||| |||  || |||||||||||| ||||||||||||
Db  13517 CTGATCTTTATGCTGTCGCACCGCGATCTTCGTCGACAGCACGACCGACGTGGTGCCCCAGATCAGC  13576

Qy   1897 gaccgctccacccgtcgagct  1919
          |||||||  |||||||||| ||
Db  13577 GAGCCGCAGTACCCCCGTCCAGTT  13599

RESULT 7
MSGB1551CS/c
LOCUS       MSGB1551CS      36548 bp    DNA     linear    BCT 17-DEC-2001
DEFINITION  Mycobacterium leprae cosmid B1551 DNA sequence.
ACCESSION   L78813
VERSION     L78813.1  GI:1377770
KEYWORDS    .
SOURCE      Mycobacterium leprae (clone: cosmid B1551) (tissue library: Lorist
            6) DNA.
ORGANISM    Mycobacterium leprae
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium.
REFERENCE   1  (sites)
  AUTHORS   Eiglmeier,K., Honore,N., Woods,S.A., Caudron,B. and Cole,S.T.
  TITLE     Use of an ordered cosmid library to deduce the genomic organization
            of Mycobacterium leprae
  JOURNAL   Mol. Microbiol. 7 (2), 197-206 (1993)
  MEDLINE   93188700
REFERENCE   2  (bases 1 to 36548)
  AUTHORS   Smith,D.R., Richterich,P., Rubenfield,M., Rice,P.W., Butler,C.,
            Lee,H.M., Kirst,S., Gundersen,K., Abendschan,K., Xu,Q., Chung,M.,
            Deloughery,C., Aldredge,T., Torrey,D., Maher,J., Lundstrom,R., Tulig,C.,
            Falls,K., Imrich,J., Torrey,D., Engelstein,M., Breton,G., Madan,D.,
            Nietupski,R., Seitz,B., Mao,J.I. et al.
  TITLE     Multiplex sequencing of 1.5 Mb of the Mycobacterium leprae genome
  JOURNAL   Genome Res. 7 (8), 802-819 (1997)
  MEDLINE   97413161
   PUBMED   9267804
COMMENT     This sequence was generated by the Genome Sequencing Center at
            Genome Therapeutics Corporation (Collaborative Research Division),
            100
            Beaver St., Waltham, MA, 02154.  Please contact Doug Smith
            (smith@cric.com) for further information.  The sequence represents
            the
            insert of a Lorist 6 cosmid clone from a mapped set of clones
            constructed from M. leprae genomic DNA isolated from armadillo
            liver
            [3].  The sequence may not represent the entire cloned insert of
            the
            cosmid if an overlapping region was previously sequenced from
```

Db   9666 ATCGAACACAGTCTGGTCCACAGATCAGCGAGCCACAGTATCCGGTACAGC 9615

RESULT 8
LOCUS   MSGB1554CS   36548 bp   DNA   linear   BCT 17-DEC-2001
MSGB1554CS/c
DEFINITION   Mycobacterium leprae cosmid B1554 DNA sequence.
ACCESSION   L78814
VERSION   L78814.1 GI:1377771
KEYWORDS   .
SOURCE   Mycobacterium leprae (clone: cosmid B1554) (tissue library: Lorist
                6) DNA.
  ORGANISM   Mycobacterium leprae
                Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
                Actinomycetales; Corynebacterineae; Mycobacteriaceae;
                Mycobacterium.
REFERENCE   1 (sites)
  AUTHORS   Eiglmeier,K., Honore,N., Woods,S.A., Caudron,B. and Cole,S.T.
  TITLE   Use of an ordered cosmid library to deduce the genomic organization
                of Mycobacterium leprae
  JOURNAL   Mol. Microbiol. 7 (2), 197-206 (1993)
  MEDLINE   93188700
REFERENCE   2 (bases 1 to 36548)
  AUTHORS   Smith,D.R., Richterich,P., Rubenfield,M., Rice,P.W., Butler,C.,
                Lee,H.M., Kirst,S., Gundersen,K., Abendschan,K., Xu,Q., Chung,M.,
                Deloughery,C., Aldredge,T., Maher,J., Lundstrom,R., Tulig,C.,
                Falls,K., Imrich,J., Torrey,D., Engelstein,M., Breton,G., Madan,D.,
                Nietupski,R., Seitz,B., Mao,J.I. et al.
  TITLE   Multiplex sequencing of 1.5 Mb of the Mycobacterium leprae genome
  JOURNAL   Genome Res. 7 (8), 802-819 (1997)
  MEDLINE   97413161
  PUBMED   9267804
COMMENT   This sequence was generated by the Genome Sequencing Center at
                Genome Therapeutics Corporation (Collaborative Research Division),
                100
                Beaver St., Waltham, MA, 02154. Please contact Doug Smith
                (smith@cric,com) for further information. The sequence represents
                the
                insert of a Lorist 6 cosmid clone from a mapped set of clones
                constructed from M. leprae genomic DNA isolated from armadillo
                liver
                [3]. The sequence may not represent the entire cloned insert of
                the
                cosmid if an overlapping region was previously sequenced from
                another
                clone. Coding sequences larger than 60 amino acids were predicted
                on
                the basis of codon usage and homology information. An attempt was
                made
                to locate the most probable start site based on codon usage,
                homology,
                the presence of a Shine-Dalgarno sequence, or overlapping orf that
                suggested translational coupling. It is possible that the actual
                start
                site differs from the one selected.
FEATURES   Location/Qualifiers
     source   1..36548
                     /organism="Mycobacterium leprae"
                     /specific_host="Dasypus novemcinctus"
                     /db_xref="taxon:1769"
                     /clone="cosmid B1554"
                     /tissue_lib="Lorist 6"
                     /note="The liver of the armadillo was used to isolate the
                     Mycobacterium leprae."
BASE COUNT   6874 a   10335 c   11630 g   7709 t
ORIGIN

Query Match   37.3%; Score 716.8; DB 1; Length 36548;
Best Local Similarity 63.3%; Pred. No. 1.2e-160;
Matches 1134; Conservative 0; Mismatches 652; Indels 6; Gaps 2;

Qy   133 gcgtttggattcaaccgctctccatcattgatattgcacactcccaccaaccactcgt 192
           || |||| |||||||||||||||||||||||||||||||||| ||||||| || ||
Db 10746 CGGCGCACGTTCACCGTTGGCTCGTTTCTGTCCGCGGTATCGACTCCACCAGCCATCGCA 10687

Qy   847 ccgcttgcaaagcgccacaacccctgacctgctcacctTcaccaccggtttcgagcgtgaa 906
           | || |||||| |||||| ||| || ||||| || || |||| |||| ||||
Db 10686 GCACTCGGGCATCCGACAACATCCGCGGCTGATTACATTTACCACCGGCTTCGAGCATGAG 10627

Qy   907 ggctactcggagtcgatgtggctgcggagtccgccgcTcgattggcggctgagcacatc 966
           || || | || ||| || | || || ||||| | |||| | |||| |||||
Db 10626 GGATTCTCCGAGATCGACGTGGCGGCGTGCTGCTTCGGCTGAAGCAATCGTGCGCCCACATC 10567

Qy   967 gtgaagattgtctcgcctgaggaataacgccaacgcgattcctaagatcatgtggtacttg 1026
           || |||| || | || |||| || ||||| || | |||| |||||||||||||||||
Db 10566 GTGAAGGTGGTTACCCCACGAGTTCGTCGCCGCCCTGCCGGAGATCGTCTGGTATCTC 10507

Qy   1027 gatgacctgctagctgacccatcattggtccgcgctgtacttcgtggcagcggaagcacgt 1086
           |||| | | || ||||| |||| | ||||| | ||| |||||||||| |||| |
Db 10506 GACGAGCCAATTGCCGACCCGGCCCTGGTTCCGCGTGTTTTTCGTCGCCGGCAAGCTCGC 10447

Qy   1087 aagcacgtcaaggtgtgctgtcggcggaggcgcagatgagctgttcgttggtggatacacc 1146
           ||| ||| ||||| || ||| || ||| | ||| || ||| | ||||| || ||||
Db 10446 AAGCACGTCAAGGTGTGTTGTCGGCCGAGGCCCACGAGCTGTTCGGCGGGGCTACAGA 10387

Qy   1147 atttacaaagagcgctatcgctgctcagttgagagatccctcgccccactacgtaaa 1206
           || ||||| || || || || || ||||| |||| || || ||||| | |
Db 10386 ATCTATCGAGAGCCCGTCGTTGAAGCCCTTCGACTACCTACCCGGGCCCACTGCCGACGC 10327

Qy   1207 ggcctgggaaagctcagcaaggttctgccagacgcgcatgaaggcaagtcccttcttgag 1266
           || |||| ||||| | ||| || | || | |||||| |||| || |||| || ||
Db 10326 TCGATGGGCAAGGTGTCCAAATCTCGCCCGAAGGTGATGCGCGGCGGCAAGAATCTGCTACAC 10267

Qy   1267 cgtggctccatgaccatggaagagcgctactacggcaacgctcgctccttcaatttcgag 1326
           ||||||| || | ||| | || |||| | || ||||| | | | || |||||| |
Db 10266 CGGCGGATCACTCACGCTCGAGGAACCCTACTACGCAATGCCCGCAATTTCTCGGAGCG 10207

Qy   1327 cagatgcaacgcgttattccatgggcaaagcgcgaatgggaccacgcggagtcactgca 1386
           |||| |||||| ||| | | || ||| |||||| | || |||||| | ||||| | |||
Db 10206 CAGTTGCAGGCTGTGCTACCCAGGTTCCACACGGGCTGGACCCCACCACGCGGACGGCA 10147

Qy   1387 ccgatcctacgcacaatcccgaacctttgatccagtagcccgcatgcaacacctggatctg 1446
           || || | || ||| |||| || ||||| || ||| | ||| || |||||||| |||
Db 10146 GCGGTGTACGGCGAATCCGTCGCGGTGGACCCGGTAGCCCGGCCATGCAGCAGCATCGATCTG 10087

Qy   1447 ttcacctggatgcgcggcgacatcctggtcaaggctgaaggtgacaagatcaacatggcggaactcc 1506
           || | |||||||| | || || ||| ||||| | || |||| || || | |||| ||||
Db 10086 TTCACCTGGCTACGTGGCGGACATTTGGCCAAGGCCGCAAGATGACAAGATGGCCAATTCG 10027

Qy   1507 cttgagctgcgagttccattcttggataaggaagtttcaaggttgcagagaccattcct 1566
           || || ||||| | || || ||| || ||||| | || ||||| | |||||| ||
Db 10026 CTGGAGCTGCGCGGTACGGTTCCTTGATCCCGAAGTGTTCCGCGGCGTCCCGGTTACCC 9967

Qy   1567 tacgatctgaagattgccaacggtaccaccaagtacgcgctgcgcaggcgcactcgagcag 1626
           ||||| | ||| ||||| | || |||||||||||| |||||| | |||| |||| ||| ||
Db   9966 ATGCAGGCAAAGATCAACATCCCGCACCACCACAGAAATACGCACTGCCGGCGGCGGCAAGTCC 9907

Qy   1627 attgttccgctcacgtttgcaccgcaagaagctggcttcctgttccctgcgccac 1686
           || | | || || ||||| | || || | |||| || | || | ||
Db   9906 ATCGTCCCGGCACACGTGTGCATCGCCCAAACTCGGGTTCCCGGTCCCAATCCGGCAT 9847

Qy   1687 tggcttgccggcgatagcgtgttcgttgggcggcaggacaccattaaggaaatccggtact 1746
           |||||| | || || ||| | ||||| ||||| |||| ||| || || |||| |
Db   9846 TGGCTGCCGGCCGAACTGCTGGTGTAAGCTCGTCGCCAGGCC 9787

Qy   1747 gaagatatctcaacaagcaggctgtgctggatatgctgaacgagcaccgcgatggcgtg 1806
           || |||| ||| ||| | || | ||| ||| | || || || |||| || ||||
Db   9786 GAACATCTGGTCAACCTGGTCAACCTAGCCGCCCGTGCACCGGATGCTTGACGAGCACCGAGGTGGCATC 9727

Qy   1807 tcagtaatcattccgctcgactgtgactgttctgtcatttatggtgtggcacggcattttt 1866
           || | ||| || | | ||| | |||| | | | ||| || || ||||||| ||| |
Db   9726 AGCGGATCATAGTCGCGGGCGTGTGACAGGGCTGTCCTAATTTTATGCGTGGCACCCGCCGGCCATCTTC 9667

Qy   1867 gtggaaaccgcattgatccacagattgaggaccgctcctacccggtcgagc 1918
           ||||| || | |||| ||| ||||| ||| | || |||| | ||

```
RESULT   9
BSUB0016/c
LOCUS       BSUB0016       215640 bp    DNA    linear    BCT 26-NOV-1997
DEFINITION  Bacillus subtilis complete genome (section 16 of 21): from 2997771
            to 3213410.
ACCESSION   Z99126
VERSION     Z99119.1  GI:2635411
KEYWORDS    .
SOURCE      Bacillus subtilis
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 215640)
  AUTHORS   Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G.,
            Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borchert,S.,
            Borriss,R., Boursier,L., Brans,A., Braun,M., Brignell,S.C.,
            Bron,S., Brouillet,S., Bruschi,C.V., Caldwell,B., Capuano,V.,
            Carter,N.M., Choi,S.K., Codani,J.J., Connerton,I.F., Cummings,N.J.,
            Daniel,R.A., Denizot,F., Devine,K.M., Dusterhoft,A., Ehrlich,S.D.,
            Emmerson,P.T., Entian,K.D., Errington,J., Fabret,C., Ferrari,E.,
            Foulger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A.,
            Galleron,N., Ghim,S.Y., Glaser,P., Goffeau,A., Golightly,E.J.,
            Grandi,G., Guiseppi,G., Guy,B.J., Haga,K., Haiech,J., Harwood,C.R.,
            Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Hullo,M.F.,
            Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y.,
            Klaerr-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P.,
```

```
          Koningstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A.,
          Lardinois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,
          Masuda,S., Mauel,C., Medigue,C., Medina,N., Mellado,R.P.,
          Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M.,
          Ogawa,K., Ogiwara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M.,
          Portetelle,D., Porwollik,S., Prescott,A.M., Presecan,E., Pujic,P.,
          Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M.,
          Rivolta,C., Rocha,E., Roche,B., Rose,M., Sadaie,Y., Sato,T.,
          Scanlan,E., Schleich,S., Schroeter,R., Scoffone,F., Sekiguchi,J.,
          Sekowska,A., Seror,S.J., Serror,P., Shin,B.S., Soldo,B.,
          Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K.,
          Takeuchi,M., Tamakoshi,A., Tanaka,T., Terpstra,P., Tognoni,A.,
          Tosato,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassarotti,A.,
          Viari,A., Wambutt,R., Wedler,E., Wedler,H., Weitzenegger,T.,
          Winters,P., Wipat,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K.,
          Yoshida,K., Yoshikawa,H.F., Zumstein,E., Yoshikawa,H. and
          Danchin,A.
  TITLE     The complete genome sequence of the gram-positive bacterium
            Bacillus subtilis
  JOURNAL   Nature 390 (6657), 249-256 (1997)
  MEDLINE   98044033
  REFERENCE 2 (bases 1 to 215640)
  AUTHORS   Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
            Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
            Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
            adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
            68 89 48
  FEATURES             Location/Qualifiers
       source          1..215640
                       /organism="Bacillus subtilis"
                       /strain="168"
                       /db_xref="taxon:1423"
       gene            complement(8..1001)
                       /gene="ytnM"
       terminator      complement(8..24)
                       /gene="ytnM"
       CDS             complement(99..1001)
                       /gene="ytnM"
                       /function="unknown"
                       /codon_start=1
                       /transl_table=11
                       /protein_id="CAB14906.1"
                       /db_xref="GI:2635412"
                       /db_xref="SPTREMBL:O34430"
                       /translation="MKKLIVFAFIGLLSQLIDGSLGMAYGVTSTSLLLAFGITPAVAS
       ASVHLAEVVTAASGVSHIKFGNVDKQTVYQLVIPGSIGAFLGAAFLSQLPGDVAKPV
       ISLFLLLGGVVLIRFLFQYKPALEKKHVPLNRKQSIPLGVIAGFADATGGGWGPVT
       TPILLSRKGLSPRKVVGTVDTSEFAIAVSATAGFLISLGWEDVNWLWVFSLMAGGIIA
       APIRAWLVQKFHPQLMGVLVGGFILVNARTLINEWIANTAVHPLIYTAIGAIHLSAV
       LFVLSKIGNRNIVKTAVDVHLKEK"
       gene            complement(1020..2270)
                       /gene="hipO"
       CDS             complement(1020..2270)
                       /gene="hipO"
                       /EC_number="3.5.1.32"
                       /codon_start=1
                       /transl_table=11
                       /product="hippurate hydrolase"
                       /protein_id="CAB14907.1"
                       /db_xref="GI:2635413"
                       /db_xref="SPTREMBL:O34980"
                       /translation="MSLDYWRNIEGSSYPYQTTGNDILTLKEESNPVNLSTLEKQLIGI
       RRHLHQYPELSKEFETTAFIKKCLKEKGIQIRPTALKTGVFADIAGESEGPAIALRA
       DIDALPIEEKTGLPYASKHKGIMHACGHDFHTRALLGAAFLLKENQDSLKGKIRLLFQ
       PAEEAGAGATKVIEDGCQLDGIDAVIGLHNKPDIAVGTVGLKTGPLMAAVDRFKVEIEG
       KGAHAALPHNGFDPIIGASQLIVALQTIVSRNVNPLQSAILTVGKINGGSTWNVIPDT
       VVIGTVRFTDSEVRNQVKQRFFAVTEQISAAFSLKANVKWHSGCPPPLCNDEAITGLV
       RDAAHKAKLQVIDPAPSTAGEDFAYYLEHIPGSFAFFGTDGDHDWHHPAFTIDETAII
       KASYFLYESAKRLLDSNEESKISD"
       gene            complement(2288..2980)
                       /gene="ribR"
       CDS             complement(2288..2980)
                       /gene="ribR"
                       /function="regulation of riboflavin biosynthesis genes"
                       /note="alternate gene name: ytnK"
                       /codon_start=1
                       /transl_table=11
                       /product="transcriptional regulator"
                       /protein_id="CAB14908.1"
                       /db_xref="GI:2635414"
                       /db_xref="SWISS-PROT:P94465"
                       /translation="MTIIAGTVVKGKQLGRKLGFPTANVDAKIHGLRNGVYGVLATVN
       HQFHLGVMNICVKPTVGSNLEKTLEIFLEFDFPHRDIYGEKIECSILFKIREERRFDSLE
       SLTKQIKKDISCVAKRPELIGIMAPNKKESLLSHQELNLPDLCFYKKCNNLYGVNRGV
       YNVIDNWFFEYGITQVAYRRIYILSFSFLKEDNPKVSSKYIRFGAGGLADKLNRFIS
       SYVEESEENILG"
       gene            complement(3027..4355)
                       /gene="ytnJ"
       CDS             complement(3027..4355)
                       /gene="ytnJ"
                       /function="unknown"
                       /note="similar to nitrilotriacetate monooxygenase"
                       /codon_start=1
                       /transl_table=11
                       /protein_id="CAB14909.1"
                       /db_xref="GI:2635415"
                       /db_xref="SPTREMBL:O34974"
                       /translation="MTRADFIQFGAMIHGVGGTTDGWRHPDVDPSASTNIEFYMKKAQ
       TAEKGLFSFIFIADGLFISEKSIPHFLNRFEPITILSALASVTKNIGLVGFTSTSFTE
       PFTISRQLMSLDHISGGRAGWNLVTSPQEGAARNHSKSNLPEHTERYEIAQEHLDVVR
       GLWNSWEHDAFIHNKKTGQFFDQAKLHRLNHKGKYFQVEGPLNIGRSKQGEPVVFQAG
       SSETGRQFAAKNADAIFTHSNSLEETKAFYADVKSRAADEGRDPSSVRIFPCISPIVA
       DTEERAEKKYREFAELIPIENAVTYLARFFDDYDLSVYPLDEPFPDIGDVGKNAFQST
       TDRIKREAKARNLTLREVAQEMAFPRTLFIGTPERVASLIETWENAEAADGFIVGSDI
       PGTLDAFVEKVIPILQERGLIRQDYRGGTLRENLGLGIPQHQSVLHSSHH"
       gene            complement(4352..4633)
                       /gene="ytnI"
       CDS             complement(4352..4633)
                       /gene="ytnI"
                       /function="unknown"
                       /codon_start=1
                       /transl_table=11
                       /protein_id="CAB14910.1"
                       /db_xref="GI:2635416"
                       /db_xref="SPTREMBL:O34639"
                       /translation="MSDVVNIVVWSKKGCSYCEEVKNYLNEKGFPFQNIDVSEKEKLR
       DILQVKYGVRHVPVVEIGRGNQYQGITEIGIEHLDLALANHAQIKEAKR"
       gene            complement(4648..5652)
                       /gene="ytmO"
       CDS             complement(4648..5652)
                       /gene="ytmO"
                       /function="unknown"
                       /note="similar to hypothetical proteins"
                       /codon_start=1
                       /transl_table=11
                       /protein_id="CAB14911.1"
                       /db_xref="GI:2635417"
                       /db_xref="SPTREMBL:O34846"
                       /translation="MIRLSILDQSLIGEGETAADTLQHTVKLAQMAEECGYHRFWVAE
       HHNNDEIAGSAPEVLLGYLAASTRKIRLASGGVMLQHYSSYKVAEQFHLLSALAPGRI
       DLGVGKAPGGFQLSTDALQAEYKRPVRQFDEKLEELTHFVRDDFPDTHRYAALRPRPQ
       VDRKPGIFLLGGSTESAISAAKLGISFVFAYFINGEEEVLKEARRAFDAHLPPGSEAE
       FHLAPAVFAAHTKEEAEKHIVSRESIKVVLKDGRKVNVGSREQAEAYLENVTEPYDII
       VQKTGIIAGTKEEVAEELTRLSGTYKINDFVIFTPIKNAVEKQLSYQLLSDAVLAAKR
       "
       gene            complement(5649..6428)
                       /gene="hisP"
       CDS             complement(5649..6428)
                       /gene="hisP"
                       /codon_start=1
                       /transl_table=11
                       /product="histidine transport protein (ATP-binding
                       protein)"
                       /protein_id="CAB14912.1"
```

```
            /db_xref="GI:2635418"
            /db_xref="SPTREMBL:O34900"
            /translation="MIEIKNIHKQFGIHHVLKGINLTVRKGEVVTIIGPSGSGKTTFL
            RCLNLLERPDEGIISIHDKVINCRFPSKKEVIHWLRKQTAMVFQQYHLFAHKTVIENVM
            EGLTIARKMRKQDAYAVAENELRKVGLQDKLNAYPSQLSGGQKQRVGIARALAIHPDV
            LLFEDEPPAALDPELVGEVLEVMLEIVKTGATMIVVTHEMEFARRVSDQVVFMDEGVIV
            EQGTPEEVFRHTKKDRTRQFLRRVSPEYLFEPKEHIKEPVI"
  gene     complement(6425. .7132)
            /gene="ytmM"
  CDS      complement(6425. .7132)
            /gene="ytmM"
            /function="unknown"
            /note="similar to amino acid ABC transporter (permease)"
            /codon_start=1
            /transl_table=11
            /protein_id="CAB14913.1"
            /db_xref="GI:2635419"
            /db_xref="SPTREMBL:O34931"
            /translation="MQFDFPFIVSAMKEMVKTIPLTLMMAVLPIVFGFLVALGNIIVR
            IFRIKGLVACSRFYVSFFRSTPAILHIMLIYLGIPPVADQVSSFFHLGWSANEIPVSM
            FVIMALSLTAGAYLTEIIRSGLAMDFGQVERAYSIGLTYSGTFRRVILPQALKVSIP
            NFTNLGIGFLHTSIAAIVAVPEITGTATIVASDNYAFLEAFIGAAIIYWVLTLILES
            ANGVLERKAARFQGGTL"
  gene     complement(7162. .7881)
            /gene="ytmL"
  CDS      complement(7162. .7881)
            /gene="ytmL"
            /function="unknown"
            /note="similar to amino acid ABC transporter (permease)"
            /codon_start=1

Query Match    24.3%;    Score 466.8;  DB 1;  Length 215640;
Best Local Similarity   55.7%;    Pred. No. 7.5e-101;
Matches  997;  Conservative   0;  Mismatches  757;  Indels  36;  Gaps   4;

Qy       88 caccggtgctgccgacgatgacgccgcacttggcatcgacgatgacgccgatgacgagcgagcatggtttgattcaac  147
            ||||| ||||| || ||||| |  || ||  ||||  |  ||| ||||  |     |  |  ||||| | ||| ||
Db   128893 CACCGCGGTCCTGACAGTGATGATGATATTCCATGATGAGCACCACCACTGCGGCTTCGGATTCAGA        128834

Qy      148 cgcctccatcattgatattgcacactcccaccaccactgcgtgggacctgcggat           207
            || || | || | |||  |||||| |||| | ||||  ||| ||| |  |  ||            --AT   128786
Db   128833 CGCCTACCGTCACTATAAAGAAGAGAAAATGGTGGACAGCCTTTACCAT

Qy      208 gaacccgaccgcacgcaatgactttcaacgtgagatctacaactacgttgagctgcgt                267
            |||| ||  |||   || | |  | |||   | || |  ||    |   |||| | ||
Db   128785 GAAGATGAAACATATTGGATTATCTTTAACGCAGAAAATCTATAACTATATCGAACTGACGA   128726

Qy      268 aaagagctctggatttggatatacctttaatacttctggcgtcggagtggcggcaattgtt                327
            || ||| ||  |||||||| ||  |  ||    ||||||| |  || ||||| | ||  ||
Db   128725 GAAGAACTTGAAGCAACTTGGGTACACAGGGTACACATTCAACACGGATTCGGACACAGAGGTTCTTCTT   128666

Qy      328 gtcggtttccaccactgggcgagtcggtcgagcatctccggaatgttcggcatt                387
            || || ||||| ||| ||| |  |  ||| | ||  | |  |  |||  ||||||
Db   128665 GCGACTACCGTCACTATAAAGAAGAGAAAATGGTGGACAGCCTTTACCGTCACTATATGCCTTTTGCCTTT   128606

Qy      388 gccattggaatacaaaggaaaagtcgctttccttgcgcgtgatcagttcggcatcaag                447
            ||||| ||||| |  ||| |  || || | ||| |  | | |  |  |   || ||
Db   128605 CTGATTGGAATAAAAAACGATCATGTGCTTATTGGGCAAGAGTCCAAGAGATCCGGCATTAAA   128546

Qy      448 cactgtctacgcaccacggagcatgcaacgtgtctcctcagacgtgtctcctcagaagagagaccatc                507
            || ||  ||  || || ||||  |  |  |  | |  | ||| |  ||| | ||| |||| || | ||
Db   128545 CCGTTGTACTACACAACGATCAATGATCAGGTTTTATTTTGCCTCAGCATCAGACGAGAGATTTTA   128486

Qy      508 ttggagatggccggagagatgaatctagatctggccttgataagcgcaccattgagcac                567
            ||||| || | || ||| ||| |  |  | ||||| |  ||   || | || ||| ||
Db   128485 ATGGTTGCTGCTACGGAATGATATTTGAAATA------GATAAAGAGGCATTGCAGCAG   128438

Qy      568 tacgtgacctgcagtctgcccgacccagatacccttcacgcgcagatttcccgcctg                627
            || || ||||| ||||| |  || || || ||| ||| | ||| ||  |  ||| | ||
Db   128437 TACATGTCTTTCCCAATTTGTTCCTGAACCAAGCACGCGCGTCTAGTGCCCATGTGAAAAAAGTA   128378

Qy      628 gagtcaggtcgcaccgccaacagttcgtccggccggcaagctgaacagcgttacttc                687
            || || ||| |||||||||| ||| | || ||| ||| |||| | || | |  |||
```

```
Db   128377 GAGCCGGTTCACAATTCACAATCCCCCCGACGGCGCGATATCACATTCAAAACGTATTTC  128318

Qy      688 aagcctcagttccagtacagaaggtctgtaaaggaggaggaggaggaccctcctcgatcgc           747
            |||||| |||||||| | |  |   ||  |||| || | |    | ||| ||  || ||| ||
Db   128317 AAAGGCCGGT---------AAGCCGGCAACTTC--------AAGCGGTCAAGACGTGAAGAA  128267

Qy      748 attgccaggtgttgagga tagcgcgagaaagcatatcgtcgcgacgtgaccgtaggc           807
            |  |  || ||  |  |  |||||   ||||| | || |||  |||| | || | | ||||  ||||
Db   128266 GTCCGAGACGACCGATCTATGACTTCTCATATGACGAAGTGACGGTTCCTGTCGGT  128207

Qy      808 tcgttcctttccgggcgcattgactcaaccgcaatgcgccttgcaaagcgccacaac           867
            || || |  ||| ||| | || |  | |  |  |  |  | ||| | ||| |  | | |
Db   128206 TCGTTCCTGTCTGCGGCTGGCGGATTGATTCTTTATTGTGTTGCGTTGTGCGGAAGGAATTTCAT  128147

Qy      868 cctgacctgctcacctcaccacggtttcgagcggtactcggaggtcggatgtg           927
            || ||||| ||| ||   || ||| ||  || |  ||| ||||| || | || ||
Db   128146 CCAAGCTTAAAAACATTCTCTGTCCGGCTTGAACACAGCAGGCTTCAGTGAGGTTGATGTA  128087

Qy      928 gctcggagtccgccgctgcgcggcgtgcatgcctgagcacatcgaagatctgtctcgcctgag           987
            || ||| | ||| |||| || ||  |  |   ||| | | |  | ||| ||  | |  ||| ||
Db   128086 GCGAAAGAAACGGCAGCGGCGCCTGGGCATTGAAAATATCAGCAAAGTCATTTCGCCCTCAG  128027

Qy      988 gaatacgccaacgcgcgattcctaagatcatagtggtatctggatgatgatccgtagctgaccca          1047
            || |  | |  ||  |  | |    |   |||  | || ||  ||| ||| |  ||  |||   |||
Db   128026 GAATACATGATGAGCTTCCGAAGATCGTATGCGCATTTGATGACCCGCCTTCTGATCCG  127967

Qy     1048 tcattggtccggctgtacttcgtgtggcagcggaagcacgtaagcacgtcaaggttgtgctg          1107
            || ||||| || | |||| |||| |||| |||  |  |  |  |||||   |||||  |  |||
Db   127966 GCAGGCCATTCCGTTGCTGTTACTTTGTTGCAAAGAACAAAAACATGTTACCGTTGCTTGCTTTA  127907

Qy     1108 tctggcgagggcgcagatgacgcgttcggtggatacacccatcacaaagaggcgcatcg          1167
            ||||| |||||| | |  || || ||||| ||||| | ||| |  || || | || ||
Db   127906 TCAGGTGAAGGAGGCGGATGAGCTCTTCGGCGGATAACATTTAACATTTACCGTGAACCGCTTTCG  127847

Qy     1168 cttgctccattgagaaggagatccctcccactacgctaaaggcctgggaaagctcagcaag          1227
            ||| || ||| ||||  |  ||  || | ||||| || || ||||| ||||| || || |||
Db   127846 CTTAAACCGTTGAACGGATTCCGCTGCCTGAAAAAAGATGCTTCGCACGTTGCTGCC  127787

Qy     1228 gttctgccagacggcgatgaagggcaagtcaatcttggggacctgcgctggtgggctccatggaa          1287
            || |  | |  ||| | || | |||||||| ||| ||| || ||| || ||  |  | ||| | |
Db   127786 GTTATGCCTGAAGGTATGAGAGGGAAGAGCCTGCTTGAGCCGTGCCACTCCGCTTCAG  127727

Qy     1288 gagcgctacggcaacgcgtcctcctcaatttcgagcagatgcagcgcgttattcca          1347
            || ||  | |||   |||  |  || | ||||  |||| | | ||||| |||| ||||
Db   127726 GATCGTTATATCGGTAACGCGAAAAATCTTTGAAGAGTCGTCAAAAAGCAGCTGTTGAAG  127667

Qy     1348 tgggcaaagcggaatggcgcagacaccgcgaagtcactacgcaacatcccgc          1407
            | ||| |||  |||  ||| ||| |||| || | ||  | |  ||| ||| |  |
Db   127666 CATTATAATCCGAACCTTCATATCGCGATGTGACGAAGACTTACTTTACAGAAAGCAGC  127607

Qy     1408 aactttgatccagtagccgcgcgcgcaacacctgatctgtcacctggatgcgcgcgac          1467
            || ||||| || ||| ||| ||| || |  ||| ||| ||   ||| ||| | ||||| ||||
Db   127606 TCGTACAGCGATATCAACACAAATGCAGTATGCTCGATATCCACACTTGGATCGCGGCGGCGGAC  127547

Qy     1468 atcctggtcaaggtgacaagatcaacatggcgaactccctcgagcctgcgagttccattc          1527
            || ||| || ||||| || |  |||| | ||||  | |  | |  |||  ||| ||| || |
Db   127546 ATTTGTTAAAGGCGGATAAAAATGACGATGGCGAATTCTTTGGAACTGGCGTGTGCCCTTTC  127487

Qy     1528 ttggataaggaagtttcaaggttgcagacgttccgagacctgagattgccaac          1587
            || |||||| |  |||||| |||| |  |    || ||| ||||| | || |  |
Db   127486 TTGGATAAAGTCGTATTCGACGTCGCGCTTCAAAAATTCCTGAGCGTGAAAAACGAAGAAC  127427

Qy     1588 ggtaccaccaagtacgcgctgcgcagggcactcgagcgaattgtctccgccgcacgtttttg          1647
            ||||||| ||||| |  || |   | || || ||||   ||| | || || | ||  ||
Db   127426 GGCACGCACGAAGTATCTTCTTCGCAAGGCCCTGAAGGAATTGTGCTCGAGCACGTATTA  127367

Qy     1648 cacccagaaagctgggcttcccatgccactgcttgccggcgatgaagctg          1707
            |||| |||||| | ||| ||| ||| |  || | |||  |   || |||  |
Db   127366 AACCCGAAAAAGCTTGGTTTTCCCGTCGTACCGATCCGGTCATTGGCT---GAAAAACGAAATG  127310

Qy     1708 ttcggtgggcgcaggacaccattaaggaatccggtactgaagagatctcaacaagcag          1767
            ||  || ||||| |||| |||| | |  ||| | |||  || ||| | || | || |
Db   127309 AACGAATGGGTGCGGAATATATTATTCAAGAAGCCAAACGGATGCTTTATATCCATAAGGAT  127250
```

```
Qy      1768  gctgtgctggatatgctgaacgagcaccgcgatggcgtgtcagatcattcccgtcgactg 1827
              |||||||||| ||| |||| ||| ||||| ||||||| ||||| ||| ||| | ||||||
Db 127249  TATGTCCTTCAGCTTCTTGAAGACCATTGTGCGGACAAGGCTGATAACAGCGCGTAAAATT 127190

Qy      1828  tggactgttcgtcattatggtgtggcacggcattttgtggaaaaccg 1877
              ||||||| || |||||| ||  |||||| |||| |||| |||||||||
Db 127189  TGGACTGTGCTAATCTTTATGATCTGGCACAGCATCAATATTGAAAACG 127140
```

RESULT 10
AF008220
LOCUS   AF008220   220060 bp   DNA   linear   BCT 04-FEB-1998
DEFINITION Bacillus subtilis rrnB-dnaB genomic region.
ACCESSION   AF008220
VERSION   AF008220.1 GI:2293135
KEYWORDS   .
SOURCE   Bacillus subtilis.
ORGANISM   Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.

REFERENCE   1 (bases 1 to 7430)
AUTHORS   Green,C.J., Stewart,G.C., Hollis,M.A., Vold,B.S. and Bott,K.F.
TITLE   Nucleotide sequence of the Bacillus subtilis ribosomal RNA operon,
          rrnB
JOURNAL   Gene 37 (1-3), 261-266 (1985)
MEDLINE   86031361
REFERENCE   2 (bases 153210 to 153762)
AUTHORS   Connors,M.J., Mason,J.M. and Setlow,P.
TITLE   Cloning and nucleotide sequencing of genes for three small,
          acid-soluble proteins from Bacillus subtilis spores
JOURNAL   J. Bacteriol. 166 (2), 417-425 (1986)
MEDLINE   86195826
REFERENCE   3 (bases 213161 to 218473)
AUTHORS   Ogasawara,N., Moriya,S., Mazza,P.G. and Yoshikawa,H.
TITLE   Nucleotide sequence and organization of dnaB gene and neighbouring
          genes on the Bacillus subtilis chromosome
JOURNAL   Nucleic Acids Res. 14 (24), 9989-9999 (1986)
MEDLINE   87117549
REFERENCE   4 (bases 200404 to 201481)
AUTHORS   Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.
TITLE   Cloning and nucleotide sequence of phoP, the regulatory gene for
          alkaline phosphatase and phosphodiesterase in Bacillus subtilis
JOURNAL   J. Bacteriol. 169 (7), 2913-2916 (1987)
MEDLINE   87250247
REFERENCE   5 (bases 201234 to 203212)
AUTHORS   Seki,T., Yoshikawa,H., Takahashi,H. and Saito,H.
TITLE   Nucleotide sequence of the Bacillus subtilis phoR gene
JOURNAL   J. Bacteriol. 170 (12), 5935-5938 (1988)
MEDLINE   89053932
REFERENCE   6 (bases 142232 to 144147)
AUTHORS   Grundy,F.J. and Henkin,T.M.
TITLE   Cloning and analysis of the Bacillus subtilis rpsD gene, encoding
          ribosomal protein S4
JOURNAL   J. Bacteriol. 172 (11), 6372-6379 (1990)
MEDLINE   91035248
REFERENCE   7 (bases 133624 to 134990)
AUTHORS   Henkin,T.M., Grundy,F.J., Nicholson,W.L. and Chambliss,G.H.
TITLE   Catabolite repression of alpha-amylase gene expression in Bacillus
          subtilis involves a trans-acting gene product homologous to the
          Escherichia coli lacI and galR repressors
JOURNAL   Mol. Microbiol. 5 (3), 575-584 (1991)
MEDLINE   91260441
REFERENCE   8 (bases 140810 to 142610)
AUTHORS   Henkin,T.M., Glass,B.L. and Grundy,F.J.
TITLE   Analysis of the Bacillus subtilis tyrS gene: conservation of a
          regulatory sequence in multiple tRNA synthetase genes
JOURNAL   J. Bacteriol. 174 (4), 1299-1306 (1992)
MEDLINE   92138624
REFERENCE   9 (bases 217570 to 220060)
AUTHORS   Putzer,H., Gendron,N. and Grunberg-Manago,M.
TITLE   Co-ordinate expression of the two threonyl-tRNA synthetase genes in
          Bacillus subtilis: control by transcriptional antitermination

                                                            involving a conserved regulatory sequence
JOURNAL   EMBO J. 11 (8), 3117-3127 (1992)
MEDLINE   92347349
REFERENCE   10 (bases 134990 to 141290)
AUTHORS   Grundy,F.J., Waters,D.A., Takova,T.Y. and Henkin,T.M.
TITLE   Identification of genes involved in utilization of acetate and
          acetoin in Bacillus subtilis
JOURNAL   Mol. Microbiol. 10 (2), 259-271 (1993)
MEDLINE   95020526
REFERENCE   11 (bases 162129 to 164080)
AUTHORS   Grundy,F.J., Waters,D.A., Allen,S.H. and Henkin,T.M.
TITLE   Regulation of the Bacillus subtilis acetate kinase gene by CcpA
JOURNAL   J. Bacteriol. 175 (22), 7348-7355 (1993)
MEDLINE   94042910
REFERENCE   12 (bases 7009 to 15526)
AUTHORS   Kiel,J.A., Boels,J.M., Beldman,G. and Venema,G.
TITLE   Glycogen in Bacillus subtilis: molecular characterization of an
          operon encoding enzymes involved in glycogen biosynthesis and
          degradation
JOURNAL   Mol. Microbiol. 11 (1), 203-218 (1994)
MEDLINE   94195107
REFERENCE   13 (bases 196487 to 200620)
AUTHORS   Jin,S. and Sonenshein,A.L.
TITLE   Identification of two distinct Bacillus subtilis citrate synthase
          genes
JOURNAL   J. Bacteriol. 176 (15), 4669-4679 (1994)
MEDLINE   94321340
REFERENCE   14 (bases 16985 to 19588)
AUTHORS   Abe,A., Koide,H., Kohno,T. and Watabe,K.
TITLE   A Bacillus subtilis spore coat polypeptide gene, cotS
JOURNAL   Microbiology 141 (Pt 6), 1433-1442 (1995)
MEDLINE   95400496
REFERENCE   15 (bases 131934 to 133970)
AUTHORS   Bolotin,A., Khazak,V., Stoynova,N., Ratmanova,K., Yomantas,Y. and
          Kozlov,Y.
TITLE   Identical amino acid sequence of the aroA(G) gene products of
          Bacillus subtilis 168 and B. subtilis Marburg strain
JOURNAL   Microbiology 141 (Pt 9), 2219-2222 (1995)
MEDLINE   96118703
REFERENCE   16 (bases 25258 to 31212)
AUTHORS   Rowland,B., Hill,K., Miller,P., Driscoll,J. and Taber,H.
TITLE   Structural organization of a Bacillus subtilis operon encoding
          menaquinone biosynthetic enzymes
JOURNAL   Gene 167 (1-2), 105-109 (1995)
MEDLINE   96144257
REFERENCE   17 (bases 196487 to 200620)
AUTHORS   Jin,S., De Jesus-Berrios,M. and Sonenshein,A.L.
TITLE   A Bacillus subtilis malate dehydrogenase gene
JOURNAL   J. Bacteriol. 178 (2), 560-563 (1996)
MEDLINE   96134995
REFERENCE   18 (bases 129888 to 132207)
AUTHORS   Varon,D., Brody,M.S. and Price,C.W.
TITLE   Bacillus subtilis operon under the dual control of the general
          stress transcription factor sigma B and the sporulation
          transcription factor sigma H
JOURNAL   Mol. Microbiol. 20 (2), 339-350 (1996)
MEDLINE   96310371
REFERENCE   19 (bases 81540 to 91690)
AUTHORS   Bower,S., Perkins,J.B., Yocum,R.R., Howitt,C.L., Rahaim,P. and
          Pero,J.
TITLE   Cloning, sequencing, and characterization of the Bacillus subtilis
          biotin biosynthetic operon
JOURNAL   J. Bacteriol. 178 (14), 4122-4130 (1996)
MEDLINE   96312354
REFERENCE   20 (bases 49093 to 51682)
AUTHORS   Yocum,R.R., Perkins,J.B., Howitt,C.L. and Pero,J.
TITLE   Cloning and characterization of the metE gene encoding
          S-adenosylmethionine synthetase from Bacillus subtilis
JOURNAL   J. Bacteriol. 178 (15), 4604-4610 (1996)
MEDLINE   96345628
REFERENCE   21 (bases 100760 to 102298)
AUTHORS   Kappes,R.M., Kempf,B. and Bremer,E.
TITLE   Three transport systems for the osmoprotectant glycine betaine

```
JOURNAL    operate in Bacillus subtilis: characterization of OpuD
           J. Bacteriol. 178 (17), 5071-5079 (1996)
MEDLINE    96359357
REFERENCE  22 (bases 1 to 220060)
AUTHORS    Lapidus,A., Galleron,N., Sorokin,A. and Ehrlich,S.D.
TITLE      Sequencing and functional annotation of the Bacillus subtilis genes
           in the 200 kb rrnB-dnaB region
JOURNAL    Microbiology 143 (Pt 11), 3431-3441 (1997)
MEDLINE    98048467
REFERENCE  23 (bases 1 to 220060)
AUTHORS    Lapidus,A., Galleron,N., Sorokin,A. and Ehrlich,D.
TITLE      Direct Submission
JOURNAL    Submitted (13-JUN-1997) Laboratoire de Genetique Microbienne, INRA,
           Domaine de Vilvert, Jouy-en-Josas cedex 78352, France
FEATURES             Location/Qualifiers
     source          1..220060
                     /organism="Bacillus subtilis"
                     /db_xref="taxon:1423"
     rRNA            485..2034
                     /product="16S ribosomal RNA"
     rRNA            2203..5129
                     /product="23S ribosomal RNA"
     rRNA            5185..5300
                     /product="5S ribosomal RNA"
     tRNA            5322..5397
                     /note="codon recognized: GUA"
                     /product="tRNA-Val"
     tRNA            5430..5504
                     /note="codon recognized: ACA"
                     /product="tRNA-Thr"
     tRNA            5543..5618
                     /note="codon recognized: AAA"
                     /product="tRNA-Lys"
     tRNA            5629..5715
                     /note="codon recognized: CUG"
                     /product="tRNA-Leu"
     tRNA            5721..5795
                     /note="codon recognized: GGC"
                     /product="tRNA-Gly"
     tRNA            5810..5895
                     /note="codon recognized: UUA"
                     /product="tRNA-Leu"
     tRNA            5905..5981
                     /note="codon recognized: CGU"
                     /product="tRNA-Arg"
     tRNA            5997..6073
                     /note="codon recognized: CCA"
                     /product="tRNA-Pro"
     tRNA            6079..6151
                     /note="codon recognized: GCA"
                     /product="tRNA-Ala"
     tRNA            6172..6248
                     /note="codon recognized: AUG"
                     /product="tRNA-Met"
     tRNA            6251..6327
                     /note="codon recognized: AUG"
                     /product="tRNA-Ile"
     tRNA            6334..6425
                     /note="codon recognized: UCA"
                     /product="tRNA-Ser"
     tRNA            6443..6519
                     /note="codon recognized: AUG"
                     /product="tRNA-Met"
     tRNA            6531..6607
                     /note="codon recognized: GAC"
                     /product="tRNA-Asp"
```

```
Query Match      24.3%;  Score 466.8;  DB 1;  Length 220060;
Best Local Similarity   55.7%;  Pred. No. 7.6e-101;
Matches  997;  Conservative   0;  Mismatches  757;  Indels   36;  Gaps   4;
```

```
Qy      88 caccggtgtcctgacgacgcggcacttggcatgacgacgccgatgcagcgtttggattcaac 147
           |||| |||||||| ||   ||  ||   || || || |||  | ||||| |  | |||||||
```

```
Db   51534 CACCGGGTCCTGACAGTGATGGATATTTCCATGATGAGCACGTCGCTTCGGATTCAGA 51593
Qy     148 cgccttccatcattgatattgcacactccaccaccactcgttggggacctgcggat 207
           |||||   |||| |  | ||||| | ||  | |   |  |  | |||| |||||| -
Db   51594 CGGCTCAGCATTATTGATGTAGAAAATGGTGGACAGCCTTTATCAT-----AT 51641
Qy     208 gaacccgaccgtacgcaatgacttcaacgtgagatctacaactacgttgagctgcgt 267
           |||  ||  | ||   || |  ||||| ||| ||  | |||| | |||| ||||||||
Db   51642 GAAGATGAAACATATTGGATTATCTTAACCGAGAAATCTATAACTATATCGAACTGAGA 51701
Qy     268 aaagagctcggatttggatataccttaatacttctggcgatggcgaccaattgtt 327
           ||| | ||| ||| ||| ||| | |  || | || ||||| ||| |||| |||| |
Db   51702 GAAGACTTGAGCAAGGGGTACACATTCAACACCGATTCGGACACAGAGTTCTTCTT 51761
Qy     328 gtcggtttccaccactgggcgagtccggtcgtgtcgagcatcccggaaatgttcggcatt 387
           ||||| ||| | |||| ||  | || ||| || |||   ||  | ||   ||||   |||
Db   51762 GCGACTTACCGTCACTATAAAGAAGAAGCGGCGTCCAAACTGCGCGGTATGTTTTGCCTTT 51821
Qy     388 gccatttgggatacaaaggaaaagtcgcttttccttgcgtgctagtcagttcggcatcaag 447
           |||  | ||||| || |  ||    ||  |   || || || |   || | ||   || ||
Db   51822 CTGATTTGGAATAAAAAACGATCATGTGCTTTATGGGCAAGATCCATTCGGCATTAAA 51881
Qy     448 ccactgttcacgcaaccaccgagcatggcaccgtgttctcctccagagagagaccatc 507
           ||  ||||| |  |||| ||||| ||| || | ||  | |||  | |||| ||| | |||
Db   51882 CCGTGTACTACACAACGATCAATGATCAGGTTATTTGCCTCAGAGAAAAAGTTTA 51941
Qy     508 ttggagatggccgagagagtgaatctagatctggccttgataagcgcaccattggcgcac 567
           ||||  | | | | |  | | ||||   ||| |  |  |      |  | |  || |||||
Db   51942 ATGGTTGCTCAGAATGATATTGAAATA---------GATAAAAGGCATTGCAGCAG 51989
Qy     568 tacgttgacctgcagtacgtgcccgagccagataccctcacgcgcagattcccgcttg 627
           ||| || ||||| ||| ||| |   |  ||| | |||| ||| | ||  ||  | |||
Db   51990 TACATGTCTTTCCAATTGTTCCTGAACCAAGCACGCTTGATGCCCATGTGAAAAAGTA 52049
Qy     628 gagtcaggcgtgcacgcaacagttgtccgggcggcaagcagacagaagagcgttacttc 687
           || |||| | | ||| ||   || || || | || |  |  | || ||| | | |||| |
Db   52050 GAGCCGGGTCACAATTCACAATCCGCCGGACGGGCATATCACATTCAAAACGTATTTC 52109
Qy     688 aagcctcagttccagtacagaaggtcgtaaagggtaaggagagcacctcttcgatcgc 747
           |||  | | |   | ||| || ||  | |  || || ||| ||| || |  ||| | ||
Db   52110 AAAGGGCAACTC-------AAGCCGGTTCAGACTGAAGAACACAAGCTTGTGAAAGAA 52160
Qy     748 attgccaggtgttggaggatagcgtcgaaaagcatatgcgtgccgacgtgaccgtaggc 807
           |||||||| |||||||| ||| | |||| |||| ||| | |  |   | || || ||||
Db   52161 GTCCGAGACGGGCATCATGACCTCTGTAAAACGTTCATATGAAGAGTGACGTTCCTGTCGGT 52220
Qy     808 tcgttccttccgggcgcattgactcaaccgcaatcggccgcttgcaaagcgccacaac 867
           || || | || ||| |||| | ||   ||   |  | |||| |||| | ||||  |||
Db   52221 TCGTTCCTGTCTGGCGGATTGATTCTTCCTTATTGTGTTGGCGTTGCGAAGGAATTTCAT 52280
Qy     868 cctgacctgctcaccttcaccaccgtttgagcgtgagcgtagcgaggtcgggtgtg 927
           ||| | ||| | ||   |||   || ||| | | | |||| |  || || ||| |||
Db   52281 CCAAGCTTAAAAACATTCTCTGTCGGCTTTGAACACCAGGGCTTCAGTGAGGTTGATGTA 52340
Qy     928 gctgcggagtccgccgctgcggattggcgctgagcacgctgaaagtgtctgcgcctgag 987
           ||| ||| | |||||| || |||||| |||| ||| | | |  | | |  |  |  |||
Db   52341 GCGAAAGAAACGCAGCCGCCGGCGGCATTGAAAATAATCAGCAAAGTCATTTCGCCTGAG 52400
Qy     988 gaatacgccaacgcgattcctaagatcatgtgtacttggatgatccgctagctgaccca 1047
           || ||  |   | |  |   |||| | || || | | ||| | ||  ||| ||||  ||
Db   52401 GAATACATGAATGAGCTTCCGAAGATCGCATTTGATGACCCGCTTGCTGATCCG 52460
Qy    1048 tcattggtcccgctgacttcgtggcagcggaagcacgtaagcacgtcaaggttgctg 1107
           || ||||||||| | ||| | || | || | |||| |  |   ||| |||| || |||
Db   52461 GCAGGCCGATTCCGTTGTACTTTGTTGTCAAAAAACATGTTACGGTTGCTTTA 52520
Qy    1108 tctggcgaggcgcagtgagctgttcggtggatacaccatttacaaagagccgtatcg 1167
           ||||| || ||  |||  | | |  |  |  | ||    || |  |  ||||  || |
Db   52521 TCAGGTGAAGGAGCGGATGAGCTCTTCGGCGGATATAACAATTTACCGTGAACCGCTTTCG 52580
Qy    1168 cttgctccatttgagaagatccctcccaccacgtaaaggctgggaaagctcagcaag 1227
           |||| | | ||||| ||| |||| | | | | ||   |||||  || | | |||||||
Db   52581 CTTAAACCGTTTGAACCGATTCCGGCTTGCAAAAGATCTTCGCACGTTGCCC 52640
```

```
AUTHORS    Glaser,P., Frangeul,L. and Rusniok,C.
TITLE      Direct Submission
JOURNAL    Submitted (06-JUN-2001) Glaser P., Institut Pasteur, Genomique des
           Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris
           Cedex 15, FRANCE
COMMENT    E-mail: pglaser@pasteur.fr
           Phone: +33 1 45 68 89 96, Fax: +33 (0)1 45 68 87 86.
FEATURES            Location/Qualifiers
     source         1..103050
                    /organism="Listeria monocytogenes"
                    /strain="EGD-e"
                    /db_xref="taxon:1639"
     gene           complement(149..2513)
                    /gene="lmo1606"
     CDS            complement(149..2500)
                    /gene="lmo1606"
                    /note="similar to DNA translocase"
                    /codon_start=1
                    /transl_table=11
                    /protein_id="CAC99684.1"
                    /db_xref="GI:16411042"
                    /translation="MGWFKDFFFGDMDEEIDTYEDTSSRKVEKKAKQRASEVLPKSNV
                    TAIKTKERTIRNERPVAYKTAPKQKHMQPVRKQMKTQMVYQYPKGEFRFPLIPDKQKS
                    QPIQPKKQPITERQVAETQPVKEETRKRPFTATDVPSPVYAFNKRPSKFEFAVTEAEE
                    LSTIQEDLTITPVDLLDSAEAETIAFDAGLNRQIEEEVVSVPVMEEVVTEQPEVEVNP
                    EPVEQQEPPARVSLITEEPAQTKTTTRSKQVESNRQEQLLKSRIPFNVMMVKDKQALQ
                    KEDAQEINVQQPVEIEAEQTNIVQQTQVATASYPMNYEFPSPSGLLHPVPVSKREDDSWL
                    QMQQEMLDETLENFNVQASVVNRTQGPAVTRFEVQPEKGVKVSKITNLJTDDIKLNLAA
                    KDIRIEAPIPGKSTVGIEIPNQTSRPVMLSELMNTEAFQTSASPLTAALGLDISGTPI
                    ITDLQKMPHGLIAGATGSGKSVCINSLLVSLLYKATPDQLKLLLIDPKMVELAPYNRI
                    PHLVSPVITDAKAATVALKWAVEEMERRYQLFSHTGVRNMEKYNEYASHPDHTGEKLP
                    YILIVIDELADLMMVAPNDVEESISRIAQKARACGIHMIVATQRPSVDVITGLIKANI
                    PTRVSFSVSSQIDSRTILDASGAEKLLGKGDMLFLPSGASKPVRLQGTFVSGENIDAV
                    VAHVRSQGEADYIFEEQELLVKETAKENTDELFEEACDFVLSQNAASTSLLQRHFRIG
                    YNRAARLMESLENHQIVSGINGSKPRDVIITKDQLAKLRNKES"
     RBS            complement(2508..2513)
                    /gene="lmo1606"
     gene           complement(2769..3427)
                    /gene="pheT"
     terminator     complement(2769..2791)
                    /gene="pheT"
     CDS            complement(2810..3427)
                    /gene="pheT"
                    /note="similar phenylalanyl-tRNA synthetase (beta
                    subunit)"
                    /codon_start=1
                    /transl_table=11
                    /protein_id="CAC99685.1"
                    /db_xref="GI:16411043"
                    /translation="MIVNAFYNEKGVGDTLLIETGEVTFENREWERRGDVARIFDRET
                    KEVAGFNIFNSSKYFDITANGKVELTEELVALMNRVIEKAGEPEIVADLSPKFVGY
                    VESKEKHPNADKLNVCKINVGDKVLQIVCGAPNIDVGQKVVAKIGAVMPSGLIIKPS
                    NLRGEDSFGMVCAARELAIPDAPTEKGIMVLEESATIGEAFPVNF"
     gene           complement(join(3433..4233,4241..4246))
                    /gene="lmo1608"
     CDS            complement(3433..4233)
                    /gene="lmo1608"
                    /note="similar to unknown proteins"
                    /codon_start=1
                    /transl_table=11
                    /protein_id="CAC99686.1"
                    /db_xref="GI:16411044"
                    /translation="MAKMTLKMKERLEKELQAPNRQFSYNRDNDALTVVQSGKKVTL
                    AIPQIIANYENDGDAAVEKIVYVEEGFRAAAGNVELENNKANVYPVVRATSFPGETK
                    AGEVLLTDDHTAETKIFYAVDLGKSYRFIEESMLKKAQLTHEIREAAFNNLANLEIP
                    LKKDSVNGEDFYFVRTNDGYDASRLNEAFLREMREKLGEWVLAVPHQDVLLIGDIQ
                    DNTGYDVLAHTMDFFADGLVPITSLPFVYNNGKLEPIFIMAKNRLKE"
     RBS            complement(3435..3440)
                    /gene="pheT"
     gene           complement(4241..4246)
                    /gene="pheT"
     RBS            complement(4241..4246)
                    /gene="lmo1608"
```

```
Qy    1228 gttctgccagacgcatgaaggcaagtcccttcttgagcgtggctccatgaccatggaa 1287
           || || |||| ||  || ||   || ||| ||||||| |||  || ||||||||| ||
Db   52641 GTTATGCCTGAAGGTATGAGAGGGAAGAGCCTGCTTGAGCGCGGCTTGCCACTCCGCTTCAG 52700

Qy    1288 gagcggctactacggcaacgctcgctcctcaattcgagcagatgcaacgcgttattcca 1347
           ||||    ||||||  |||||||    ||  |||| |||| ||||||    |  ||||
Db   52701 GATCGTTATATCGGTAACGCGGAAAATCTTTGAAGAGTCTGTCAAAAGCAGCTGTGTGAAG 52760

Qy    1348 tgggcaaagcgcgaatggaccaccgaccgcgaagtcactgcaccgatctacgcacaatcccgc 1407
           |  |||||| || ||  || ||  || ||||| ||||  ||   ||| || ||||||||| ||
Db   52761 CATTATAATCCGAACCTTTCATATCCGCATGTGACGATGGACAGCTTACTTTACAGAAACAGAGC 52820

Qy    1408 aactttgatccagtagcccgcatgcaacacctggatctgttcacctggatgcggcggac 1467
           |||| |||   ||||| |||||  || || |||| |||  | |||||||||||||||||||
Db   52821 TCGTACAGCGGATATCAACAAAATGCAGTATGTCGGATATCCACACTTGGATGCGGCGGCGAC 52880

Qy    1468 atcctggtcaaggctgacaagatcaacatgcgaactccctgagctgcgagttccattc 1527
           ||||  ||| ||  || | ||||| ||||| | |||| |||||||| |||||||  ||||
Db   52881 ATTTGTTAAAGGCGGATAAAAATGACGATGCGGATAAAATCTTTGGAACTGCGTGTGCCCGTTC 52940

Qy    1528 ttggataaggagtttttcaaggttgcagagaccattccttacgatctgaagattgccaac 1587
           ||||||| || |||| ||| | || ||  | |  || ||||||||||| |||||| |||
Db   52941 TTGGATAAAGTCGTATTCGACGTCGCTTCAAAAATTCCTGATGAGCTGAAAAACGAAGAAC 53000

Qy    1588 ggtaccaccaagtacggcgctgcgcaggcactcggagcagattgttccgcctcacgtttttg 1647
           || |||||||| |||||||||| |||||  ||| ||| ||||| | ||||||||||||  |
Db   53001 GGCACGCAGCAAGTATCTTCTTCGCAAAGCCCCTGAAGGAATTGTGCCTGAGCACGTATTA 53060

Qy    1648 caccgccaagaagctgggcttccctgttcccatgcgcgccactggcttgccgcgatgagctg 1707
           ||  || | |||||| ||||||||||||||| |||||| ||| |||||| || ||||| |||
Db   53061 AACCGCCAAAAAGCTTGGTTTCCCTGTTCCCTGACCGATCCGCATTGGCT---GAAAAACGAAATG 53117

Qy    1708 ttcggttgggcgcaggacaccattaaggaatacggtactgaagatatcttcaacaagcag 1767
           || |||| |||| ||| ||||||| |  |||||| ||  |||| ||||||| ||||| ||
Db   53118 AACGAATGGGGTGCGGAATATATTATTCCAGAAAGCCAAACGGATCCGTTATATCCATAAGGAT 53177

Qy    1768 gctgtcggatatgctgaacgagcaccgcgatggcgtgtcagatcattcccgtcgactg 1827
           ||  || | |||| ||| |||| || |||| | |||| || ||  ||||||||| ||||
Db   53178 TATGTCCTTCAGCTTCTTGAAGACCATTGTCGCGACAAGGCTGAATAACAGCGCGTAAAATT 53237

Qy    1828 tggactgttctgtcattatggtggtcacggcatttttgtggaaaaaccg 1877
           ||||||| || ||  || ||  |||||||||||    ||||||||||||
Db   53238 TGGACTGTGTGTAATCTTTATGATCTGGCACAGCATCAATATTGAAAAACG 53287
```

```
RESULT 11
LOCUS       AL591980/c
DEFINITION  AL591980              103050 bp    DNA     linear    BCT 04-DEC-2001
            Listeria monocytogenes strain EGD, complete genome, segment 8/12.
ACCESSION   AL591980 AL591824
VERSION     AL591980.1 GI:16411041
KEYWORDS    .
SOURCE      Listeria monocytogenes.
ORGANISM    Listeria monocytogenes
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Listeria.
REFERENCE   1 (sites)
AUTHORS     Glaser,P., Frangeul,L., Buchrieser,C., Rusniok,C., Amend,A.,
            Baquero,F., Berche,P., Bloecker,H., Brandt,P., Chakraborty,T.,
            Charbit,A., Chetouani,F., Couve,E., de Daruvar,A., Dehoux,P.,
            Domann,E., Dominguez-Bernal,G., Duchaud,E., Durant,L.,
            Dussurget,O., Entian,K.D., Fsihi,H., Portillo,F.G., Garrido,P.,
            Gautier,L., Goebel,W., Gomez-Lopez,N., Hain,T., Hauf,J.,
            Jackson,D., Jones,L.M., Kaerst,U., Kreft,J., Kuhn,M., Kunst,F.,
            Kurapkat,G., Madueno,E., Maitournam,A., Vicente,J.M., Ng,E.,
            Nedjari,H., Nordsiek,G., Novella,S., de Pablos,B., Perez-Diaz,J.C.,
            Purcell,R., Remmel,B., Rose,M., Schlueter,T., Simoes,N.,
            Tierrez,A., Vazquez-Boland,J.A., Voss,H., Wehland,J. and Cossart,P.
TITLE       Comparative genomics of Listeria species
JOURNAL     Science 294 (5543), 849-852 (2001)
MEDLINE     21557279
REFERENCE   2 (bases 1 to 103050)
```

```
gene            complement(4268..4593)
                /gene="lmo1609"
CDS             complement(4268..4593)
                /gene="lmo1609"
                /note="similar to thioredoxin"
                /codon_start=1
                /transl_table=11
                /protein_id="CAC99687.1"
                /db_xref="GI:16411045"
                /translation="MKNLESVEQFNAIKSEGKSVFMFSADWCGDCKFIEPVMPEIEAE
                NEDFSFYHVDRDEFIDLCADLAIFGIPSFLAFEDGEEVGRFVSKDRKTKEEINDFLAA
                I"
RBS             complement(4588..4593)
                /gene="lmo1609"
gene            complement(4691..4909)
                /gene="lmo1610"
CDS             complement(4691..4909)
                /gene="lmo1610"
                /note="hypothetical gene"
                /codon_start=1
                /transl_table=11
                /protein_id="CAC99688.1"
                /db_xref="GI:16411046"
                /translation="MHKIMKKPFVDGMLLVASLVFLLGYATSMPYFRDSEIGWIWT
                TLIAGIITLFFTFFNDFLEKKARSKVR"
gene            complement(join(4902..5975,5985..5991))
                /gene="lmo1611"
CDS             complement(4902..5975)
                /gene="lmo1611"
                /note="similar to aminopeptidase"
                /codon_start=1
                /transl_table=11
                /protein_id="CAC99689.1"
                /db_xref="GI:16411047"
                /translation="MEKETLAMFKELTELQGTSGDEHRIRAFMRKELEPVSDEIIQDG
                LGGIFGVRHGAEEGPRVLVAAHMDEVGFMVTQITDNGLLRFQTIGGWNPQVLQAQRVQ
                IMTPIGPVIGVVASVPPHLLSEAERSKPTDPKNLLIDIGADDKADAEKIGIKPGQFIV
                PVAEFTPLANPKKILAKAWDNRYGVGLAIELMKELKGESLPNTLFAGANVQEEVGLRG
                AGVSANMIKPDLFFALDASPANDTTGNKSQFGSQIGQGFLLRIFDRTMIMHRGMREFLL
                DTAETNKIPYQYFVSPGTDAGKVHTSLSGIPSAVIGVPARYIHSSNSILHVDDYAAA
                KEMITTLIRGLDKTTYETIKNNA"
RBS             complement(5985..5991)
                /gene="lmo1611"
gene            join(6160..6484,6494..6515)
                /gene="lmo1612"
RBS             6160..6165
                /gene="lmo1612"
RBS             6173..6484
                /gene="lmo1612"
CDS             complement...
                /gene="lmo1612"
                /note="similar to unknown proteins"
                /codon_start=1
                /transl_table=11
                /protein_id="CAC99690.1"
                /db_xref="GI:16411048"
                /translation="MNWKAFIAGVGAGAAAGHLVYHYLLSDKTISGDVILEKVKDAFK
                QEGPIEGSWIQLKKQHYKKFAIDTFVYHGGITCIREGEKKQFEIADANTGTIIDVYL
                A"
gene            complement(join(6494..6515,6521..6769,6776..6781))
                /gene="lmo1613"
terminator      6494..6515
                /gene="lmo1613"
terminator      complement(6494..6515)
                /gene="lmo1613"
CDS             complement(6521..6769)
                /gene="lmo1613"
                /note="similar to unknown proteins"
                /codon_start=1
                /transl_table=11
                /protein_id="CAC99691.1"
                /db_xref="GI:16411049"
                /translation="MDIEQKKALVTRFAKIEGHVRSIKTMTEEERDLETIMQQIAAVK
                KAMDAAAKLIYTEQMKELIEQGETDEAVIKKKIDSFIR"
RBS             complement(6776..6781)
```

```
terminator      complement(6885..6902)
                /gene="lmo1613"
gene            complement(6885..7768)
                /gene="lmo1614"
CDS             complement(6903..7754)
                /gene="lmo1614"
                /note="similar to unknown proteins"
                /codon_start=1
                /transl_table=11
                /protein_id="CAC99692.1"
                /db_xref="GI:16411050"
                /translation="MDSIQIGKIKIYWLRGGYTHFDGGAMFGVVPKPLWEKKYPVNDK
                NQLANVTDPMFFQYLGKNYLIDSGLGNRRLTEKQKRNYGVTEESFVLEDLAQLSIAPE
                DIDYVLMTHLHFDHVLGLTGISEEGNDYSIFPNAEIWTSETEWEEMQHPNIRSKATYW
                QENWLRIEGQIHTFTKEKEFPNDAIKMTHTGHSNGHSVIWLESDGEKAIHMADIFPTF
                AHQNVLWVTAYDDYPMTSISAKQAIFKQTFGANYWFLSYHDSRFRAVQIAENGDVKAE
                LKIEHRN"
RBS             complement(7763..7768)
                /gene="lmo1614"
gene            complement(7783..8459)
                /gene="lmo1615"
terminator      complement(7783..7805)
                /gene="lmo1615"
CDS             complement(7815..8459)
                /gene="lmo1615"
                /note="similar to unknown proteins"
                /codon_start=1
                /transl_table=11
```

```
Query Match     20.3%;    Score 389.2;   DB 1;   Length 103050;
Best Local Similarity 53.4%;    Pred. No. 2.6e-82;
Matches  954;   Conservative  0;   Mismatches 798;   Indels  36;   Gaps  5;

Qy      88  caccggtggtcctgacgatgccggcactggcatgacgccgatgacgcgttggatccaac  147
            ||| ||  ||  |||  || ||  ||  ||||| ||  || ||  || |||| ||  | |
Db   65006  CACCGTGGGCCCAGATGATGAAGGATACTTCACAGATGATCACGTGCAGTTTCGTTTCCGC  64947

Qy     148  cgcctcccatcattgatattgcacactcccaccaacacgcgtggggacctgcggat  207
            ||||| |  ||  ||||| |  ||  ||     |||   || |||| ||        |
Db   64946  CGTTTAAGTATTATTGATGTAGAAAACGGTCATCAACCGTT----------AACGTAT  64899

Qy     208  gaaccgaccgtacgcaatgactttcaacgtgagatcacaactacgtgagctcgt  267
            || | |  |  || |||  |||  |||   ||  ||| || || |||  |  ||||
Db   64898  GAAAATGAACGTTACTGACTTATTTTAATGGAGAGTTTATAACTATGTCGAACTCCGC  64839

Qy     268  aaagagctcggatttgggatatacctttaatacttctggcgatggcgagccaattgtt  327
            || |||||||||  || ||  |  ||  |||  |   ||||| ||  ||  ||| | |
Db   64838  GAGCAATTAGTTGCTCGAAGGAATGACATTTGAAACCGAAAGTGATACCGAAGTAATTATC  64779

Qy     328  gtcggttccacactgggcgagtccgtgtcgagcatcccgaatgttcgcgcatt  387
            || ||||||| ||| ||  ||  ||  || || | ||   || | ||  |||||
Db   64778  GCTACATATGCGAAATACAAAGAAAAACAGCAGAACGCCTTCGCGGGATCGTTTGGTTTT  64719

Qy     388  gccatttgggatacaaaggaaaagtcgctttccttgcgcgtgatcagttcggcatcaag  447
            ||||| |||||  | || |||||| |  |  | || |   | |  | | ||  ||| |||
Db   64718  GTTATTTGGGACCAAACCAAGAAGAAGAACTTGTTGTTCCTGTTTACGGCGGCCACCGACCGACCCATTCGGCATTAAA  64659

Qy     448  ccactgttctacgcaccgagcatggcaccgtgttctcctcagaagaagaagaccatc  507
            ||||| ||||||    |||| ||  ||  ||  |||||  |  ||||||| |  |||
Db   64658  CCATTTTCTATGCAGAGAGACGGACAAGCTGTTCATGGGTTCTGAAAAGAAATCCATT  64599

Qy     508  ttggagatggccgagagaatctagatctgggcctgataagcgcaccattgagcac  567
            ||||| ||||| |||  | ||  ||      ||  || ||  || || ||| ||||
Db   64598  TTACATGCCGTTAAA----------AGAAAAGAATTGGACGAAGTGTCATTACAAAAC  64551

Qy     568  tacgtggacctgcagtacgtgtccgagccagataccccttcacggcgagatttcccgcttg  627
            ||  || ||  ||||  |   |  ||  ||||  |   |  | |||  || | |   || |
Db   64550  TATATGCATTACCAATTGTTCCAGAACCAGATTCCTTAACTAAAAATGTCAAACGTTTA  64491

Qy     628  gagtcaggctgcaccgcaacagttcgccaacagttcggggcggcaagctggacagaagcgttacttc  687
            || |  || |  || || ||||  |  |||       ||  ||  |||  ||  |  | || |  |||
Db   64490  GAGCCAGGACACATCAGTTACCAACGATGGAAATTACGACTTATTGG  64431
```

```
Qy  1768 gctgtgctggatatgctgaacgagcaccgcgatggcgtgtcagatcattcccgtcgactg 1827
         ||| ||  |||  ||| ||  ||||||||||| || |||| |||| |||| |||| ||||
Db 63362 TACGTGCTAGACTTACTACAGACGATCACTGTGCCGGTAAATTCGACTACAGTCGTAAAATC 63303

Qy  1828 tggactgttctgtcatttatgtgtggcacggcattttgtggaaaac 1875
         ||| || ||  || ||| ||| ||||| ||||||||||||
Db 63302 TGGACAGTCGTTATTTCATGATTTTCATGATTTTCATGATTGGTACGATGTATACGTAGCAGAC 63255
```

RESULT 12
AL596169/c

LOCUS   AL596169   239050 bp   DNA   linear   BCT 04-DEC-2001
DEFINITION Listeria innocua Clip11262 complete genome, segment 7/12.
ACCESSION   AL596169 AL59202
VERSION   AL596169.1 GI:16414035
KEYWORDS   .
SOURCE   Listeria innocua
  ORGANISM Listeria innocua
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Listeria.
REFERENCE   1 (sites)
  AUTHORS   Glaser,P., Frangeul,L., Buchrieser,C., Rusniok,C., Amend,A.,
            Baquero,F., Berche,P., Bloecker,H., Brandt,P., Chakraborty,T.,
            Charbit,A., Chetouani,F., Couve,E., de Daruvar,A., Dehoux,P.,
            Domann,E., Dominguez-Bernal,G., Duchaud,E., Durant,L.,
            Dussurget,O., Entian,K.D., Fsihi,H., Portillo,F.G., Garrido,P.,
            Gautier,L., Goebel,W., Gomez-Lopez,N., Hain,T., Hauf,J.,
            Jackson,D., Jones,L.M., Kaerst,U., Kreft,J., Kuhn,M., Kunst,F.,
            Kurapkat,G., Madueno,E., Maitournam,A., Vicente,J.M., Ng,E.,
            Nedjari,H., Nordsiek,G., Novella,S., de Pablos,B., Perez-Diaz,J.C.,
            Purcell,R., Remmel,B., Rose,M., Schlueter,T., Simoes,N.,
            Tierrez,A., Vazquez-Boland,J.A., Voss,H., Wehland,J. and Cossart,P.
  TITLE   Comparative genomics of Listeria species
  JOURNAL   Science 294 (5543), 849-852 (2001)
  MEDLINE   21537279
REFERENCE   2 (bases 1 to 239050)
  AUTHORS   Glaser,P., Frangeul,L. and Rusniok,C.
  TITLE   Direct Submission
  JOURNAL   Submitted (09-JUL-2001) Glaser P., Institut Pasteur, Genomique des
            Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris
            Cedex 15, FRANCE
COMMENT   E-mail: pglaser@pasteur.fr
            Phone: +33 (0)1 45 68 89 96, Fax: +33 (0)1 45 68 87 86.
FEATURES   Location/Qualifiers
     source   1..239050
                     /organism="Listeria innocua"
                     /strain="Clip11262"
                     /db_xref="taxon:1642"
     gene   complement(57..1355)
                     /gene="lin1549"
     CDS   complement(57..1340)
                     /gene="lin1549"
                     /note="similar to unknown protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC96780.1"
                     /db_xref="GI:16414036"
                     /translation="MAIQPLAYRMRPKKALDEIVGQTHLVGKDKIIYRMVKAKQLSSMI
                     LYGPPGIGKTSIASAIAGSTKYAFRILNAVTNNKKDMEVVAAEAKMSGTVILLDEVH
                     RLDKAKQDFLPLLESGAIILJGATTSNPYIAINPAIRSRTQIFELKPLTVEDIMITM
                     DRALLDKERGLGNYEVEIDEFAKKHFATASNGDVRSALNALELAVISSEPNEDGITHI
                     TLDVAEECLQKKSLAHDKDGDAHYDVLSAFQKSVRGSDVNAALHYMGRLIEAGDLVSI
                     SRRMLVMAYEDIGLANPQAGAHTLAAIQTAEKVGFPEARIPLANAVIELCLSPKSNSA
                     IMAIDAALADIRQGNSGEVPDHLRDGHYSGAKKLGRAMDYKYPHNYDNAWVDQQYLPD
                     RLKNKLYYDPKFTSKFEQTIAGVYQKINENKAKKE"
     RBS   complement(1350..1355)
                     /gene="lin1549"
     gene   1490..1938
                     /gene="lin1550"
     CDS   1490..1912
                     /gene="lin1550"
                     /note="similar to unknown protein"
                     /codon_start=1

```
Qy   688 aagcctcagttccagtacagaggtcgtaaagggtaaggagcaggagacctcttcgatcgc 747
          ||       ||||| ||  |||| || |||    ||| ||||| ||| ||  |   | ||
Db 64430 AA------AGCATCTTTTGCACCAGTAACTAAATCAGAAGACGCATGGATTAAAGAGAGTGC 64376

Qy   748 attgccaggtgttgaggagtagcgctcgaaaagcatatggtcgcgacgtgacgtaggc 807
          |||   |||||| |||||| |||| ||    | |||| || | || ||||||  |||
Db 64375 GTG----ACGTGATGTATGATTCTGTTAAAATGCACATGCGCCTCTGATGTGCCAGTTGGT 64320

Qy   808 tcgttccttccggcggcattgactcaaccgcaattgcgccgcttgcaaagcgccacaac 867
          |  |||| |  ||| | ||| || ||  | || ||| | | ||| || |||| | |||
Db 64319 TCATTCCTATCTGGCCGTATGTGATTCATCGATTGTGCAGCAATTGCCAAAGAATACCAC 64260

Qy   868 cctgacctgctcacctcaccacggtttcgagcgtgaagctactcgaggtcgatgtg 927
          |      || |||  |  ||| ||||| ||| || |||||  | |||| |||||||
Db 64259 CCAGCAATCAAAACATTCTCAGTTGGTTTTGAACGCGATGGTTTCAGTGAAATCGATGTA 64200

Qy   928 gctgcggagtccgcgctcgagttggccgctgagcacatcgtgaagagttgtctcgcctgag 987
          ||  | |||  || |  | ||| | || ||| ||||| | || ||| ||  |||| || ||
Db 64199 GCAAAGAAACAGCCGACAAACTTGCGCTAGAAAACATTAGTTATGTTATTCTCACCAGAA 64140

Qy   988 gaatacgccaacgcgattcctaagatacatgtacttggatgatcctgtctagctgaccca 1047
          |||| |   |  |||||||| ||  || || |||||  |||||  | || | |||||||
Db 64139 GAATATGAAGAAGACTACCGCCAAAATTGTTTGGCACATGATGATCCACTTGCTGACCCG 64080

Qy  1048 tcattggtccgctgacttcgttgtggcagcggaagcacgtaagcacgtcaaggttgtgctg 1107
          ||| ||| ||| |  || |  |  | ||| ||||||| |||||  ||  | ||   | || |
Db 64079 GCTGCGTATTCCGCGTTTATTTCTTATCAGAGAGCCGCAAACAAGTAACCGTAGCATTA 64020

Qy  1108 tctgcgagggcgcgatgagctgttcggtggatgagtacaccattacaaagagccgctatcg 1167
          || | ||| || |  ||||| ||||||||| |||| | |||||||  ||||| || || |
Db 64019 TCAGGTGAAGCGCAGATGAGAACTTTTCGGTGGTTATAACATTTATAACGAACCAAATTCC 63960

Qy  1168 cttgctccatttgagagatccctcccacacgtaaagcctgggaaagctcagcaag 1227
          |||||  |  |||| || || || || | || |||| | ||| |||| ||||| |
Db 63959 CTTGCAATGTTTAACAAAATGCCAGGCGTATTTAAATCAATGTAAATAACTTGCGCGT 63900

Qy  1228 gttctgcagacggcagcaagtccatgaaggccagtcgctggctccatgaccatggaa 1287
          ||  || ||| ||  |    | |||||||| ||| ||||| || ||| ||||||||
Db 63899 ATTATGCCAGAAGGAATGCCTGGTAAAAGTTCTTAGAACGTGGAACAACCCGATGGAA 63840

Qy  1288 gagcgctactacggcaacgctcgctccttcaatttcgagcagatgcaacgcgttattcca 1347
          ||||| ||| | ||||||| | || || || |||| || ||||| ||||| |||||  ||
Db 63839 GAACGTTATATCGGAAACGCCAAAATGTTCACGGAAAAAGAAAACAAATTTTACTTCCA 63780

Qy  1348 tgggcaaagcgcgaatggaccaccgcgaagtcactgcgcacatcccgc 1407
          || |||||||| ||||||||||||||| ||| |||| | ||||||
Db 63779 AAATATCAAGCTGGTCATGACTATACGAATATTACAGAAACCATTTTATGCAGAAACAACG 63720

Qy  1408 aacttgatccagtagccgctcgctgctctggagaaaaagagatctggcatcgtgttcacctggatgcgcgacgat 1467
          ||  || | |||| ||||  ||||| |  |   | |||| ||||| ||  |  || ||  || | |||| |  ||
Db 63719 GAACGTTATATCGGAAACGCCAAAATGTTCACGGAAAAAGAAAACAAATTTTACTTCCA 63660

Qy  1468 atccggtcaaggctgacaagatcaacatggcgaactccctgagctgcgcagttccattc 1527
          |  | | |||||  |   |  |  |||| | |  ||| ||  || | |||  | |||||
Db 63659 ATTCTTCAAAGCTGACCTAAAACTAGGATTCCCCAGTACCAATTCCAGTACCAATTGCCAC 63600

Qy  1528 ttggataaggaagtttcaggtgtgcagagaccaattccttacgatctgaagattgccaac 1587
          || || |||||||| | ||| |||||| ||| |||| |  ||||| |||| || ||||||
Db 63599 CTTGATAAAGGAGTGTACAACGTGCAAGAAAATATTCCAGATACAATGAAAACAACTAAT 63540

Qy  1588 ggtaccaccagtacggcgctgcgcagggcactcgacagattgttccgccctcacgttttg 1647
          || || || ||| || || ||   |||| ||| || | || || || ||||| || ||||
Db 63539 GGAACTACTAAATATATATCTTACGTAAAGCGGCAACATTGTGTACCAGAACATGTACTT 63480

Qy  1648 cacccgaagagctggcttccctgttcccatggcgcactggcttgcgcggcgatgagctg 1707
          ||  | ||| ||| | || ||  | ||| ||||| |||| || | ||| ||||| |||
Db 63479 AATCCGCCGTAAACTAGGATTCCCAGTACCAATTCGTCACTGGTT---AAAAGACGAAATG 63423

Qy  1708 ttcggttgggcgcaggacaccattaaggataccggtactgaagatatcttcaacaagcag 1767
          ||| || ||||| ||| ||| |||  | ||| | || |  ||||| |||  ||  ||| |
Db 63422 AACGGCTTGGGTTAAAAATATCATTAAAGAATCACCAGCCGATCATTAATTAACAAACAA 63363
```

```
                      /transl_table=11
                      /protein_id="CAC96781.1"
                      /db_xref="GI:16414037"
                      /translation="MKITTKGRYGLTITLELAKRIGDGPISLRSIAQDKNLSEHYLEQ
                      LIGPLRNAGIYKSIRGAHGGYYLNGDPEKITAGDIIRTLEGPIVLVESMEDEEAAQRE
                      LWTRMRNAVRDVLDQTTLSDLLKHSTDSELTDGYMFYI"
terminator            1921..1938
RBS                   /gene="lin1550"
                      2144..2149
gene                  /gene="lin1551"
                      2144..3362
                      /gene="lin1551"
CDS                   2157..3362
                      /gene="lin1551"
                      /note="similar to ammonium transporter NrgA"
                      /codon_start=1
                      /transl_table=11
                      /protein_id="CAC96782.1"
                      /db_xref="GI:16414038"
                      /translation="MESVFMFCTLLVWLMTPGIALFYGGMVRRKNVLSTAMYSFSSM
                      AIISILWVIVGYSLAFAPGNGFIGSFDWTFLHNVGFAANDTYSDAIPHILFMMFQMTF
                      AILTVAIISGAFAERMNFSAYLIFIILWSLLVYSPVAHWVWGDGGWLRELGALDFAGG
                      NVVHISSGVTGLVLAIMIGRRKEADSASPHNLPLALIGGILVWFGWYCGFNVGSALTID
                      NVAMTAFVNTNTPAAAAGIIGWGLVEWLINKKPTMLGTISGAIAGLVSITPAAGFVTVP
                      SSLIIGFLGGALCFWAVFWLKGKVKYDDALDAFGLHGIGGIWGGIATGLFATTKVNEA
                      GADGLFYGNASLVVKQLIAIGSTVAYVAVVTALIVVVIKLFLPIRVNEEQEYKGLDLT
                      LHGEKAYQE"
gene                  3376..3741
                      /gene="lin1552"
CDS                   3376..3741
                      /gene="lin1552"
                      /note="similar to nitrogen regulatory PII protein"
                      /codon_start=1
                      /transl_table=11
                      /protein_id="CAC96783.1"
                      /db_xref="GI:16414039"
                      /translation="MSGLTKIEIITRPNRFHLFQKELAKIGVSGLTVTKALGTGLEKG
                      FIELYRGTKKESNVHERMKIEIVSTVPVEDVLRVVKETLRTGEPGDGKVFIYPLAEV
                      VKISTGETGIDALQDKPAK"
RBS                   3872..3877
                      /gene="lin1553"
gene                  join(3872..4271,4279..4298)
                      /gene="lin1553"
CDS                   3891..4271
                      /gene="lin1553"
                      /codon_start=1
                      /transl_table=11
                      /protein_id="CAC96784.1"
                      /db_xref="GI:16414040"
                      /translation="MRKIPIFTLLAAILLFGGYYLYQHASSKVDVTFNFVVDQKDKNN
terminator            4279..4298
                      /gene="lin1553"
terminator            complement(4279..4298)
                      /gene="aspS"
gene                  complement(join(4279..4298,4307..6082))
                      /gene="aspS"
CDS                   complement(4307..6082)
                      /gene="aspS"
                      /codon_start=1
                      /transl_table=11
                      /product="aspartyl-tRNA synthetase"
                      /protein_id="CAC96785.1"
                      /db_xref="GI:16414041"
                      /translation="MEKRTSYCCGELNETHIGQSVILHGWQKRRDLGGLIFIDLRDRE
                      GIVQAVVNPEFSKEALEIADSVRNEFVVTIKGVHARGEKAINDKLATGKVEVLAEEI
                      TLNTSKTPPFYIEDGVNVSDELRLKYRLYLDLRRPEMNNIFKMRHTVTRTFRNKLDAL
                      GFFDIETPYLRKSTPEGARDYLVPSRVYPGNFYALPQSPQILKQLLMTAGFDKYYQIV
                      RCFRDEDLRGDRQPEFTQIDLETSFLTKEEIQAITEDMLVDVVKEAKNITIEKPFPRM
                      TYKEAMDRFGSDKPDIRFGLELQNVSEVVKDVDFKVFQSAIENGGEVKAINAKAAATN
                      FSRKDLDALGVFVANYGAKGLAWLKIEAGELKGPIAKFFPEDKAAELQAALQAEDGDL
                      LLFAADKADIVAASLGALRNKLGKDLDLINEDELAFLWVTDWPLFEYDEEAGRYVSAH
```

```
                      HPFTLPKEEDIPFLETDSSKVMAEAYDIVLNGYEIGGGSLRIYKKEVQESMFRALGFT
                      DESAKEQFGFLMEALEYGTPPHGGIALGLDRIVMLAGRNNLRDTIAFPKTGSAVDPL
                      TNAPGEVSAAQLAELKLETVVKKETN"
gene                  complement(6084..7361)
                      /gene="hisS"
CDS                   complement(6084..7361)
                      /gene="hisS"
                      /codon_start=1
                      /transl_table=11
                      /product="histidyl-tRNA synthetase"
                      /protein_id="CAC96786.1"
                      /db_xref="GI:16414042"
                      /translation="MDLQLPRGTRDILPEEVSKWHFLESAFNKVCENYQYEIRTPIF
                      EHTELFERGVDSTDIVSKEMYTFLDKGGRSLTLRPEGTASVVRAFVEHKLYGEVSQP
                      IKMYYNEPMRYERPQGGRQRQFTQMGIEALGSDDPSIDVEVISLAMEFFRKIGLTNI
                      KLVINSLGDKESRLKHREALVAHFEPHIDEFCAEECQVRLHKNPLRLIDCKKDHDNPLI
                      QSAPSILDFLNEESVAYFENVKKYLTALEIPFEIDPTMVRGLDYNHTTFEIMSVEEG
                      FGAKTTLCGGGRYHGLVKEFGGPDTPGICRFGIGVERILLALEKAEINIPETKPLEYYV
                      ITAQPEAELKAVTLVNKLRQNGISAEKDYLKRKLKAQLKDANRKKAIYTVILGEEELQ
                      TGNYQLKNMETGEQEAVSETTIIEKLTNTKEEK"
gene                  join(7803..9086,9092..9130)
                      /gene="lin1556"
                      7803..9086
CDS                   7803..9086
                      /gene="lin1556"
                      /note="similar to N-acetylmuramoyl-L-alanine amidase"
                      /codon_start=1
                      /transl_table=11
                      /protein_id="CAC96787.1"
                      /db_xref="GI:16414043"
                      /translation="MKNKFIFITVVSILLIAAGIFTIAMANANSVVVKAEVLNVRSG
                      PGLAYDVTSQARKNEVLRVVGEENQWYKVQLDNGNSGWVASWLVENTDVSAASNSIAI
                      VSSDGGLNVREKPSTSSTSLGLLNNGDQVTVTSQQNGWAQIQYNGKSAWVSSQYLTIR
                      ESVTKVDESELQTVTIRDSTNIRNKPRGDGAVIEKANSGQFAIQGVQGDWYKIRTT
                      SGEEGVANWVDVSDKGQTSPRSKTTKLSEATIVIDPGHGGNDPGAKGANGTIEKE
                      MTLKTAKKLKEKLESRGAKVILTRNSDKYVSLKSRTNVAAENKADVFISHFDSLEDS
                      SKGVSGQTTYYYDNSDKSLAESINTTLGNDLPTTNRGSRVGDYYVVRENSQPAVLLEL
                      GYLSSAKDERRINSASYRSQIADSVVDGLSNYFSN"
gene                  complement(join(9092..9130,9135..9587,9597..9602))
                      /gene="lin1557"
terminator            complement(9597..9602)
                      /gene="lin1557"
terminator            9092..9130
                      /gene="lin1556"
CDS                   complement(9135..9587)
                      /gene="lin1557"
                      /note="similar to unknown proteins"
                      /codon_start=1
                      /transl_table=11
                      /protein_id="CAC96788.1"
                      /db_xref="GI:16414044"
                      /translation="MRVLLQRCYEASVRVEGEIISEIAGGLCLLVGFTHSDTEETVEY
                      MAKKIIGLRVFEDESEKMNISLAERGAILSVSQFTLYADVSRGKRPSFTKSAPGEKA
                      EGLNLFNNKLSDAGFIVETGVFGAFMDVKIVNNGPVTIMLDSEEMRK"
RBS                   complement(9597..9602)
                      /gene="lin1557"
gene                  complement(9603..11834)
                      /gene="relA"
CDS                   complement(9603..11819)
                      /gene="relA"
                      /note="similar to (p)ppGpp synthetase"
                      /codon_start=1
                      /transl_table=11
                      /protein_id="CAC96789.1"
                      /db_xref="GI:16414045"
                      /translation="MAKEQNLTAEQVIDMASHYMNQEHLALVKKAYEFARDSHKEQFR
```

```
Query Match     20.18;   Score 386.6;   DB 1;   Length 239050;
Best Local Similarity 53.3%;   Pred. No. 1.1e-81;
Matches  952;   Conservative  0;   Mismatches  799;   Indels  36;   Gaps  5;

Qy      88  caccgtggtcctgacgatgccggcacttggcatgacgaccgatgacgcagcgttggattcaac  147
            |||||||| || |||||| ||||| ||||| ||  |||| |||||| ||||  |  ||| ||||| |||
Db  219817  CACCGTGGGCCAGATGATGAAGGATACTTCACAGATGATCACGTGCGCAGTTTGGTTTTCCGC  219758
```

```
Qy    148  cgcctcccattgatattgcacactcccaccaccactgcgttgggacctgcggat  207
           | |  || ||||  |  ||||  ||| | || ||| || |  ||||||
Db 219757  CGTTTAAGTATTATTGATGTAGAAACGGTCATCACCGTT---------AACGTAT  219710

Qy    208  gaaccgaccgctacgcaatgactttcaacgtgagatctacacactgcgttgagctgcgt  267
           ||||| |||||||   |  ||  ||| |||||  ||||  || | ||| |||
Db 219709  GAAATGAACGTTACTGATTATTTTAATGGAGAAGTTATAACTATGTGGAACTCCGC  219650

Qy    268  aaagagctctcggatttgggatatacctttaatacttctggcgatggcgagccaattgtt  327
           | |  ||||||||||| | ||| || | |  || |  ||||| ||
Db 219649  GAGCAATTAGTTGCTGAAGGAATGACATTTGAAACCGAAAGTGATACAGAGTAATTATC  219590

Qy    328  gtcggtttcaccactgggcggcgagtccggtcgagcatctccgggaatgttcggcatt  387
           || ||  ||||  ||  ||  | ||  ||| |||||   ||||  |  | ||
Db 219589  GCTACATGGCGAAATACAAAGAAAAAACAGCAGAACGTCTTCGTGGGATGTTGGTTTT  219530

Qy    388  gccatttgggatacaaaggaaagtcgctttccttgcggtgatcagttcggcatcaag  447
           |||||||||||| ||| | | ||   ||| | |||  | |  ||  |||
Db 219529  GTTATTTGGGATAAACAAGAGAACTTGTTTACGGCCGCACGCGACCCCATTCGGCATTAAA  219470

Qy    448  ccactgttctacgcaaccaccgagcatggcaccgtgttctcctccagagaagagaccatc  507
           ||| | |  || ||| |  || ||| |  ||| ||||  |||| | ||||| |||
Db 219469  CCATTTTTTCTCATGGCAGAAGACGGTAAGCTATTTTATGGGTTCTGAAAAGAAATCCATT  219410

Qy    508  ttgggagtggcgaggagatgaatctagatctgggcttgataaagcgcaccattggcgac  567
           ||  ||  | ||||||| |  ||| | |||| | ||| | |||| | |||  |||
Db 219409  TTACATGCTTTAAAAGAAAAAGAACTA---------GATGAAGTTTCATTACAAAAAC  219362

Qy    568  tacgttgaccgcagtacgtgtcgccgagccagataccctcacgcgcagattcccgcttg  627
           |||||  |  ||| ||  || ||| |||| |  |||| | | |||
Db 219361  TACATGACTTATCAATTCGTTCCAGAACCAGATTCCTTAACTAAAAATGTAAAACGCTTA  219302

Qy    628  gagtcaggctgcaccgcaacagttcgtccggcggcaagctggaacagaagcgttacttc  687
           ||   || | | | |||| | ||| |  ||||| || |  |||  | ||   |
Db 219301  GAACCAGGACATCAGTTCACGAAGAAATTTGTCAACCAATGGAAATTACTTACTTATTGG  219242

Qy    688  aagccctcagttcccagtacagaaggtcgtaaagggtaaggagcaggaaccctcttcgatcgc  747
           | ||  ||  | |  ||| || |  ||  ||||||  ||  |||  ||  |||
Db 219241  AA------AGCATCTTTTGCACCAGTAACTAAATCAGAAGACCCATGGATTAAAGAAGTGC  219187

Qy    748  attgccaggtgttggaggatagcgtcgaaaagaacatatgcgtgccgacgtgaccgtaggc  807
           | || |  |||  |||  |  ||  | |||| | ||| ||||| | | ||   ||
Db 219186  GTG---ACGTAATGTGATGATTCTCGTAAAAATGCATATGCGCGTCCCGATGTGACCAGTTGGT  219131

Qy    808  tcgttccttccggggccattgactcaaccgcaattgcgccgcttgcaaagcgccacaac  867
           || ||  || |  || ||| | ||| | |||  ||  |  |  | |  |  |||||
Db 219130  TCATTCCTATCTGGCCGTATTGATTCATCGATTGATTCATCGGCGATTCCCAAAGAATACCAT  219071

Qy    868  cctgacctgctcacctcaccaccggtttcgagcgtgaaggctactcgaggtgatgtg  927
           ||| |  | | | ||  || | |||| ||||  | || ||| || |||| ||
Db 219070  CCTGCAATTAAAACATTTTCGTGTTGGTTTGAACCGCATGGTTTCAGTGAAATCGATGTA  219011

Qy    928  gctgcggagtccgccgctgattggcgctgagcaacatctgaagatgtctcgcctgag  987
           ||| || | ||||| |  | ||| | | |  |||  |||| || | | ||||  |||
Db 219010  GCAAAAGAAACAGCAGATAAGCTTGGTGTGTAGAAAAATATTAGTTAGTGTTTATCTCACCGGAA  218951

Qy    988  gaatacgccaacgcgattcctaagatcatgtggatcatgtacttggatgatctcgtagctgaccca  1047
           |||| | | | | | | |||| |  || |  ||  |  ||| | | || | ||| | |||
Db 218950  GAATATATGGTGAAAGAATTGTTTGGCATATGGCATGATGACCCACTTGCTGTGATCCG  218891

Qy   1048  tcattggtccggtacttcgtggcgagcgagcacgtaagcacgtcaaggttgtctg  1107
           || ||| || | | ||  |||  || | | ||| | | |||  | |||| | ||| |
Db 218890  GCTATTCCGCTATTTTCTTATTCAAGAGAAGCGCGTAAACAGTAACAGTAGCATTA  218831

Qy   1108  tctggcgaggcgcgatgaagcttcggtgttcggttggatacacccatttacaaagacgccgctatcg  1167
           | || || |  ||| | | |||  ||  ||||  | | | |||  ||   || |  ||
Db 218830  TCAGGTGAGGGCGCCAGCAGTGAACTTTTTGGCGGTTAAATCAATTTATTAACGGAACCAAAATTCC  218771

Qy   1168  cttgctccatttgagaagatccctcgcccccactacgtaaaggctggaaagtcagcaag  1227
           ||||| |   ||||  | |  | |  |  | |||| ||  ||| || |  ||| | ||||
Db 218770  CTTGCGCATGTGTTAACAAATGCCAGCGCTATTTAAATCAATGTTAAATATGGCGTGGCGT  218711
```

```
Qy   1228  gttctgcagacgcgcatgaagggcagtccctccttcctgagcgtggctggctccatgaccatggaa  1287
           || |||||| | | || ||| | |||||| ||  ||| ||| |  | |  || || |||||
Db 218710  ATTATGCCAGAAGGAATGCGTTGGTGGTAAAAGTTTCTTAGAAGCGGAACACCACCGATGGAA  218651

Qy   1288  gagcgctacgcaacgctcgctcctccaattcggcgagcagatgcaacgcggttattcca  1347
           |  | | || ||||  |||  |  |||||| | |||| |||| | | ||  | ||
Db 218650  GAACGTTCACTCGGAAACGCCAAAATGTTCACAGAAAAAGAAAAACAAATTTTACTTCCG  218591

Qy   1348  tgggcaaagcgaatggaccccgagagtcactcagcaccatccgcacacatcccgc  1407
           || ||||| ||| || ||  |  | ||| ||| |  | | |||| | | |  ||
Db 218590  AAATACCAATCTGTCATGATTATACGAACATTACGGATCCATTTTATGCAGAAACTAAG  218531

Qy   1408  aacttgatccagtagcccgcatgcaacacctgatctgttcacctggatcgcggcggcgac  1467
           | ||  || |  | | | | ||| ||   |||| | | |  | |||  || |  |  ||
Db 218530  AATTATCATCCCTGAGAAAGAATTGCAGTAGTCATATTGATATTCATACGTGGCTACGTGGAGAT  218471

Qy   1468  atcctggtcaaggctgacaagatcaacatggcgaactccctgagctgcagttccattc  1527
           || |  | |||||  || || || | |  |||| | || | |||  |  ||  || |
Db 218470  ATTCTCCTAAAAGCTGACGATGACGTGGCTAATTCACTGGAGTTCGTGTACCTTTC  218411

Qy   1528  ttggataaggaagtttcaaggttgcagagaccattcttccttctagatctgaagattgccaac  1587
           ||  | |||| |||   |  ||  || || ||| || ||| | | || |  | |||| | ||
Db 218410  CTTGATAAAGAAGTGTACAAATGTAGCAAGAAATATTCCAGATACAATGAAAACAACCAAT  218351

Qy   1588  ggtaccaccaagtacgcgctgcgcagggcactcgagacgatgttcgcctcacgtttg  1647
           ||  || | || |||| |||  | | | ||| | || |  ||||  ||| || | |||
Db 218350  GGAACAACCAAATACATCTTACGGCCACTTTTGTACCAGAACACGTACTT  218291

Qy   1648  caccgcaagaagctggcttccctgtcccatgcgcagcggccactggctgcggcgtgagctg  1707
           | |||  || || ||  |  ||   |  | |||||  | || | | ||| |||||||
Db 218290  AATCGCCGTAAACTAGGATTCCCAGTACCAATTCGTCACTGGCT--AAAAGACGAAATG  218234

Qy   1708  ttcggttggcgcaggacaccattaaggaatccggtactgaagatatctttcaacaagcag  1767
           ||   ||| || ||| || | | |  ||| |||| | || |  | |||   ||
Db 218233  AACGGCTTGGGTTAAAAATAATCATTGGAAATACCACCGATCATTCATTTAATTAACAAACAA  218174

Qy   1768  gctgtgctggatatgctgaacgagcagcaccgcgatggcgtcagatcattccgtcgactg  1827
           ||| | || | |  || ||| |  | | |  || | |||| |   | |  || ||||  |
Db 218173  TACGTTCCTAGGCTTACTAGACGATCACTGTGCCGGTAAATTCGATTACAGTCGTAAAATC  218114

Qy   1828  tggactgttctgtcattatggtgtggcacgcgcatttttgtggaaa  1874
           || ||| || |  | |  || | | ||   | |||||  ||| |||
Db 218113  TGGACAGTTGTTGTTATCTTTATGATTTGGTATGATGTATACGTAGCAGA  218067
```

RESULT 13
MLEPRTN4/c

LOCUS	MLEPRTN4 348450 bp DNA linear BCT 20-FEB-2001
DEFINITION	Mycobacterium leprae strain TN complete genome; segment 4/10.
ACCESSION	AL583920 AL450380
VERSION	AL583920.1 GI:13092922
KEYWORDS	.
SOURCE	Mycobacterium leprae.
ORGANISM	Mycobacterium leprae
	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
	Actinomycetales; Corynebacterineae; Mycobacteriaceae;
	Mycobacterium.

REFERENCE   1 (bases 1 to 348450)
AUTHORS   Cole,S.T., Eiglmeier,K., Parkhill,J., James,K.D., Thomson,N.R.,
             Wheeler,P.R., Honore,N., Ganier,T., Churcher,C., Harris,D.,
             Mungall,K., Basham,D., Brown,D., Chillingworth,T., Connor,R.,
             Davies,R.M., Devlin,K., Duthoy,S., Feltwell,T., Fraser,A.,
             Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Lacroix,C.,
             Maclean,J., Moule,S., Murphy,L., Oliver,Quail,M.A.,
             Rajandream,M.-A., Rutherford,K.M., Rutter,S., Seeger,K., Simon,S.,
             Simmonds,M., Skelton,J., Squares,R., Squares,S., Stevens,K.,
             Taylor,K., Whitehead,S., Woodward,J.R. and Barrell,B.G.
TITLE   Massive gene decay in the leprosy bacillus
JOURNAL   Nature 409 (6823), 1007-1011 (2001)
MEDLINE   21128732
REFERENCE   2 (bases 1 to 348450)
AUTHORS   Parkhill,J.
TITLE   Direct Submission

```
JOURNAL      Submitted (20-FEB-2001) Submitted on behalf of the Mycobacterium
             leprae sequencing teams, The Sanger Centre, Wellcome Trust Genome
             Campus, Hinxton, Cambridge, CB10 1SA, Unitie de Genetique
             Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Rouxx,
             75724, Paris Cedex, France. E-mail: parkhill@sanger.ac.uk
COMMENT      Notes:
             Details of M. leprae sequencing at the Sanger Centre are available
             from http://www.sanger.ac.uk/Projects/M_leprae/ A relational
             database containing the M. leprae sequences is available from
             http://genolist.pasteur.fr/Leproma/.
FEATURES             Location/Qualifiers
     source          1..348450
                     /organism="Mycobacterium leprae"
                     /strain="TN"
                     /db_xref="taxon:1769"
     gene            551..1786
                     /gene="ML0842"
     CDS             551..1786
                     /gene="ML0842"
                     /note="Similar to Mycobacterium tuberculosis hypothetical
                     44.6 kDa protein Rv1464 or MTV007.11 TR:O53155
                     (EMBL:AL021184) (417 aa) fasta scores: E(): 0, 43.6% id in
                     408 aa and to many other bacterial nifS-homologues, e.g.
                     Bacillus subtilis YurW protein yurW TR:O32164
                     (EMBL:Z99120) (406 aa) fasta scores: E(): 0, 46.7% id in
                     405 aa. Previously sequenced as TR:O32975 (EMBL:Z98741).
                     Contains Pfam match to entry PF00266 aminotran_5,
                     Aminotransferases class-V.
                     Similar to ML0117, ML0596 and ML1708"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAC31223.1"
                     /db_xref="GI:13092923"
                     /translation="MVDVFDIQAVRADFPILQETVNGKPLIWFDNAATTQKPQVVIDR
                     LSYFYAHENSNIHRAAHELAARATDAYEEARETARRFIGAAKAQEIIFVRGTTEAINL
                     VAYAWGGKHLQPCDEVVITHLEHHANIVPWQLLSSQTGAILKVAPVDDAGNLLMSEFE
                     DLLGPRTKLVAATQVSNALGTVTQGEKIVELGHRYGARVLIDGAQSIPHLPINVSELG
                     ADFFVFSGHKIYGPTGIGVLYGCEDVLTEMPPWQGGGNMIVDVTLERSLYQGPPNKFE
                     AGTGNIADAVGLGEALRYVERVGVQRIASHEQALLDYATPRLADIPGVRLVGTATEKA
                     SVLSFVLAGHEPLEVGKALNAEGIAVRAGHHCAQPVLRRLGLEATVRPSFAFYNTYEE
                     IDVFINVVRRIAEGGTNIG"
     misc_feature    653..1738
                     /gene="ML0842"
                     /note="Pfam match to entry PF00266 aminotran_5,
                     Aminotransferases class-V, score 23.70, E-value 1.7e-08"
     gene            1908..2395
                     /gene="lppP"
                     /note="lppP"
     CDS             1908..2395
                     /gene="ML0843"
                     /note="Possible pseudogene of M. tuberculosis orthologue
                     lppP (Best blastx score 294)"
                     /codon_start=1
                     /pseudo
                     /transl_table=11
                     /product="lipoprotein (pseudogene)"
     RBS             2549..2555
                     /note="possible RBS"
     gene            2559..4112
                     /gene="narK"
                     /note="narK"
     CDS             2559..4112
                     /gene="narK"
                     /note="Similar to Mycobacterium tuberculosis putative
                     nitrite extrusion protein narK1 or Rv2329c or MTCY3G12.05
                     TR:P71883 (EMBL:Z79702) (515 aa) fasta scores: E(): 0,
                     69.3% id in 488 aa and to Escherichia coli nitrite
                     extrusion protein 2 narU SW:NARU_ECOLI (P37758; P77696)
                     (462 aa) fasta scores: E(): 0, 36.6% id in 459 aa. There
                     is a frameshift near the C-terminus relative to the M.
                     tuberculosis homologue. Previously sequenced as TR:O32974
                     (EMBL:Z98741). Contains hydrophobic, probable
```

```
                     membrane-spanning regions."
                     /codon_start=1
                     /transl_table=11
                     /product="putative nitrite extrusion protein"
                     /protein_id="CAC31225.1"
                     /db_xref="GI:13092924"
                     /translation="MDQLVLLQAEESLYGPGRRSKTVPIDHRHSLKSSFFRLRLRRAC
                     RISHWDPEDQAAWEAGNKTIARLNLLWSVLTVHLGYSVWTLWPVWELFMPKDVYGFSA
                     GDKFLLATTALVGACLRVPYSLATALFGGRNWAIFSVMVLLIPTIATMVLLAHPGLP
                     LWPYLACAALTGLGGGNFAASMTNANAFYPHRLKGAAFGLAGGAGNLGVSIQVVGML
                     VIASVGDRKPYIVCGLYVVLLIIAGIGAMLFMNDIEHHRIGVNTIRPVLFVVSTRDS
                     WVLALLYLASFGSFIGFSFAFGQVLETNVAGGQSTAQAALHAAELAFIGPTLAAVAR
                     FWGGRLADRLGGSRVTLVVFGAMVFAAGLLGVLGIIEGSRVCPIRGVMMASYFAGFIT
                     LFILSLGNGSVYKMPTIFEACSHSLGINDDECRDWSRVISGVVIGFVAEVGALGGV
                     GIDALRESYLNTGGVTAAFWIFMLCYAAAGVLTWKMYVCRPLPGNLHDEAANAFAAS
                     VGASRTHRG"
     gene            complement(4100..4681)
                     /gene="ML0845"
     CDS             complement(4100..4681)
                     /gene="ML0845"
                     /note="Unknown function. Similar to part of some acyl-CoA
                     oxidases e.g. Arabidopsis thaliana acyl-CoA oxidase cx2
                     TR:O65201 (EMBL:AF057043) (692 aa) fasta scores: E():
                     1.5e-06, 35.2% id in 125 aa. Previously sequenced as
                     TR:O32973 (EMBL:Z98741) (193 aa) fasta scores: E(): 0,
                     99.5% id in 193 aa."
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAC31226.1"
                     /db_xref="GI:13092925"
                     /translation="MPLRPTNAASERVTKRTATETILQRMVGARQDSEKKGSLFNRGT
                     QGTIFEKCEDCLLSAVARRRQVKSKEMFAFDTFNTVQDHLLHAANEHTDRVVLETFLA
                     GIESCNPAARKLLDIVCDLYALSVVARHGTSSIDYSPPSALKAIAKGSRPMPSAAAL
                     CRDVGRRFSESQSSCVTPRCRTSRICPSELQPR"
     gene            complement(4959..5480)
                     /gene="ML0846"
     CDS             complement(4959..5480)
                     /gene="ML0846"
                     /note="Possible pseudogene similar to M. tuberculosis
                     paralogue Rv1747 (Best blastx score 127)"
                     /codon_start=1
                     /pseudo
                     /transl_table=11
                     /product="ABC transporter (pseudogene)"
     gene            complement(5522..5848)
                     /gene="ML0847"
     CDS             complement(5522..5848)
                     /gene="ML0847"
                     /note="Possible pseudogene of M. tuberculosis orthologue
                     Rv2327 (Best blastx score 184)"
                     /codon_start=1
                     /pseudo
                     /transl_table=11
                     /product="conserved hypothetical protein (pseudogene)"
     gene            5906..8080
                     /gene="ML0848"
     CDS             5906..8080
                     /gene="ML0848"
                     /note="Similar to Mycobacterium tuberculosis hypothetical
                     ABC transporter ATP-binding protein Rv2326c or MTCY3G12.08
                     SW:YN26_MYCTU (P71886) (697 aa) fasta scores: E(): 0,
                     76.9% id in 697 aa. Shares similar domains with many
                     ABC-type transporters e.g. Streptomyces roseofulvus ATPase
                     component of putative ABC transporter frnD TR:O68910
                     (EMBL:AF058302) (524 aa) fasta scores: E(): 8.4e-14, 31.5%
                     id in 495 aa and Synechococcus sp. nitrate transport
                     ATP-binding protein NrtD nrtD SW:NRTD_SYNP7 (P38046) (274
                     aa) fasta scores: E(): 1.6e-12, 34.8% id in 198 aa.
                     Previously sequenced as TR:O32971 (EMBL:Z98741). Contains
                     hydrophobic, possible membrane-spanning regions. Contains
                     2 Pfam matches to entry PF00005 ABC_tran, ABC transporter.
                     Contains 2 x PS00017 ATP/GTP-binding site motif A
```

```
                (P-loop). Contains 2 x PS00211 ABC transporters family
                signature."
                /codon_start=1
                /transl_table=11
                /product="ABC transporter"
                /protein_id="CAC31229.1"
                /db_xref="GI:13092926"
                /translation="MDLCIHRRMALLCQATRRCHRWKTQLGRCAAPDLAPGTGTIGKV
                TAHDPRHLHHRSGSLQPGELAQASVLAALCAVTAISVSVVPFAAGLALLGTVPMGLLA
                YRYRFRVLMTAMVAAGVIAFLITGLGFIAVVNSAYIGGLTGVVKRKGQTLTVIALA
                LFAGLAFGANVVALVVLGLRLHILFKAMTANVDGIAATLTWMHLPWVAVQLKRYFAD
                GLQHWPWMLLGYFVTILVVSLIGWWLSRVLERIRDIPDVHKLDAPSACNEDAPVGP
                VPWLDKVRFRYPHAGQDALREVSLDLRVGEHVAVTGANGSGKTTLMLILAGREPTSG
                TVDRPGAVGLGKLGTAVVLQHPESQVLGTRVADDVVWGLPPGTDVDVNRLLREVGLD
                AFAERDTGSLSGGELQRLALAAALAREPSLLIADEVTSMVDRQGRDALLGVLSGLTKR
                HPIALVHITHYNEADTADRTINLSDSPDNAGMAETVAPPVSTVAVDHRPHAPVLELV
                GVGHEYGSGTPWARAALHDISFVVRQGDGVLVYGSNGSGKSTLAWIMAGLMVPTTGAC
                LIDGRPTHEHVGAVALSFQARLQLMRSRVDLEVASAAGPSRPDDRRVAALGVVGLD
                PALAKRRIDQLSGGQMRRVVLAGLLACSPRALILDEPLAGLDAVSQRGLLRLLEDLRC
                ERGLTVVVTSHDFVGLEDVCPRTVHLRNGALESVTTAGTS"

misc_feature   6821. .7336
                /gene="ML0848"
                /note="Pfam match to entry PF00005 ABC_tran, ABC
                transporter, score 112.30, E-value 9e-30".
misc_feature   6842. .6865
                /gene="ML0848"
misc_feature   7109. .7153
                /note="PS00017 ATP/GTP-binding site motif A (P-loop)"

Query Match         12.8%;  Score 245.2;  DB 1;  Length 348450;
Best Local Similarity 63.3%;  Pred. No. 7.4e-48;
Matches 1134;  Conservative 0;  Mismatches 652;  Indels 6;  Gaps 2;

Qy     133  gcgtttggattcaaccgcctctccatcattgatattgcacactcccaccaccactgcgt  192
            |||||||  |||||||||||| | |||||||| | ||||| | |||| | ||||| |||||
Db   39402  GCGTTTGGTTTAACCGGTTCCATCATCGACATTGCACATCGCACATCCGGCGTCGCGG  39343

Qy     193  tgggggacctgcgggatgaaaccgaccgctacgcaatgactttcaacgtggatcctacaac  252
            ||||||||||| |||    |||| ||||||||||| | |||||||| | ||||| |||
Db   39342  TGGGGACCACCGGAGACACCCAACCCGGCTACATCCTGGTATTCAACGGCGAGATCTACAAC  39283

Qy     253  tacgttgagctgcgtaaagagc---tctcggatttgggatataccctttaatacttctggc  309
            || |||||||||| ||||   |   | |||  ||||| | |||  |  || || | |||
Db   39282  TACCTGGAGCTGCGACGAGCTGTCATCCGGCCACGCGGTCATCCGGCCTTCGCCACGACGGA  39223

Qy     310  gatggcgagccaattgttcggttccaccaactgggcggcggagatggcagtcgtggtcagcatctc  369
            |||||||| | |  | ||  |||| ||  || ||||||||  || || ||||||| |||| |
Db   39222  TACCTCGGAGCTGCGCACGACGTGTCATCCGGGATCAAACCACTGTTCCATTACTGGGCACACGGCTG  39163

Qy     370  cgcggaatgtcggcattgcattgggatacaaaggagaaaagtcgcttttccttgcgcgt  429
            ||||| | ||| |  ||| || | |  |     | ||  |  |  ||||  || ||  ||
Db   39162  CGTGGCATGTCGCCGCTACGTGGCGACATTTGCGCCAGAATTGTTCTGTGCCCGA  39103

Qy     430  gatcagttcggcatcaagccactgttctacgcaaccgagcagttgcaccgtgttctcc  489
            |  |||| ||||  |||||  ||| || ||  |  | |  || || || || || ||
Db   39102  GACCCTTCGGAGCTGCGGGATCAAACCACTGTTCATGGCGACCAAGCCACCGGTGGCC  39043

Qy     490  tcagaggaagagaccatcttggagatggccgaggagaatctagatctgggcccttgat  549
            || ||| | ||   ||| ||||| |||| |||||  | |  || || || || |||
Db   39042  AGCGAAAAGAAATGCCTGCTTGAACTGCCCGAAGTGATCAGATTGCGACACGCAGATTGAC  38983

Qy     550  aagcgaccattgagacctgacactgagacctggacgtacgtcagtacgctacgccctcac  609
            |||||  ||| |||| |||  || ||| | ||||  | ||  ||  ||| | |  |  ||
Db   38982  ACCCGCGCACTACAGACACTACACGGTCCTGCTCGAGTACGTGCCCGAGCGCGAGCGCCAT  38923

Qy     610  gcgcagatttcccggcttggagtcaggctgcgacccgcaacagttcgtccgggcgg---caag  666
            | ||||| |  |  | ||||| || ||| |||||| ||| || || |||   |  |
Db   38922  CGGCGGGTGCGCACCGTGACTGCGTGTCATCGATCCGGTATCCGTCCCGAACAGCTCGAT  38863

Qy     667  ctgaacagaaagcgttacttcaagcctcagtcccagtacagaaggtcgtaaagggtaag  726
            || | | | | || | |||  |||  | ||| || || | || |  || ||  |||
Db   38862  CCAGCCACCACCGCTACTTCGTCGCCGCAATTTGCCGCCGATCACCGCCACCAGCAGCACT  38803
```

```
Qy     727  gagcaggacctcttcgatcgcattgcccaggtgttggggatagcgtcgaaaagcatatg  786
Db   38802  GAACAGACCCGCTACGGCAGTCAGTCAGTCGGACTCAGTCGCCAAGCACACATG  38743

Qy     787  cgtgccgacgtgaccgtaggctcgttccttccggcggcattgactcaacccgcaatgcg  846
Db   38742  CGCGCCGACGACGTCACCGTTGGCTCGTTTCGTCGGCGGTATCGACTCCACAGCCATCGCA  38683

Qy     847  ccgcttgcaaggcgccacaaccctgacctgctcaccttcaacgttcggttcggagcgtgaa  906
Db   38682  GCACTGGCGATCCGACACAACATCCGCGGGTGATTACATTTACCACCGGCTTCGAGCATGAG  38623

Qy     907  ggctactcggagtcgatgtggccgagtccgccgctgcggattggcgttggcgctgagcacatc  966
Db   38622  GGATTCTCCGAGATCGACGACGACGTGCTCTTCGGCTGAAGCAATCGGTGCCGCGCCACATC  38563

Qy     967  gtgaaagattgtccgcctgagaatacgccaacgcgatcctaagatcatggtgagttg  1026
Db   38562  GTCAAGGTGGTTACCCCCAACGAGTTCGTCGCCGCCCTGCCGGAGATCGTCTCGGTATCTC  38503

Qy    1027  gatgatcctgtagctgaccatcattggtcccgctgacttcgtgcagcgggaagcacgt  1086
Db   38502  GACGAGCCAATGCCGACCCGGCCCTGGTTCCCGCTGGTTTCCGGCCGGCGAAGCTCGC  38443

Qy    1087  aagcacgtcaaggttgctgctgctgcgaggcgcgcagatgagctgtcggtggatacacc  1146
Db   38442  AAGCACGTCAAGGTGGTTGTTGTCCGGCAGCGGCCGCGAGCTGTTCGGGCGTACAGA  38383

Qy    1147  attacaaagagccgctatcgcttgctccattgagaagatccctccccactacgtaaa  1206
Db   38382  ATCTATCGGAGAGCCCGCTGTCGTTGAAGCCTTCGACTACCTACCCGGCCACTGCCGACGC  38323

Qy    1207  ggcctggaaagctcagcaagttctgccagagacgacggcatgagaggcaagtcccttcttgag  1266
Db   38322  TCGATGGGCAAGGTGTCCAAATCTCTGCCCGGAAGGTATGCGCCGCAAGAAATCTGCTACAAC  38263

Qy    1267  cgtggctccatgaccatggaagagcggctactacggcaaacgctcgctcctcattcgag  1326
Db   38262  CGCGGGATCACTCACGCTCGGACGACTACTACGCGACGGCCGAGATCTACAAAC  38203

Qy    1327  cagatgcaacgcgttattccatggcaaagcgcgaatggacaccgcgaagtcactgca  1386
Db   38202  CAGTTGCAGGCTGTGCTACCCAGGTTCACCAGGTTCCACACGGGCTGACCCACCGACGTCACGGCA  38143

Qy    1387  ccgatcctacgcacaatcccgacacttgatccagtagcccgatgcaacacctggatccg  1446
Db   38142  GCGGTGTACGCCCAATCCGCGCTGGGACCCGGTAGCCGCCGTAGCCACCATCGATCTG  38083

Qy    1447  ttcacctggatgcgcggcgacatcctgtcaagctgacaagatcaacatggccgaactcc  1506
Db   38082  TTCACCTGGCTACTGGCGACATTTGGCCAAGGCCGGCAAGAAGATGACAATGGCCAATTCG  38023

Qy    1507  cttgagctgcggagttccattcttggataaggaagtttcaaggttgcagagaccattcct  1566
Db   38022  CTTGGAGCTGCGCGGGTACCCGTTCCTTGATCCCGAAGTGTTCGCCGTGGCGTCCCGGTTACCC  37963

Qy    1567  tacgagatctgaagagaccaagtaccaccaagaccgctcgcgcaggcgcactcgagcag  1626
Db   37962  ATGCAGGCAAAAGTACCACCCGCACCCACGAAATACCGCCACTGGGGCGGCCACTGGAATCC  37903

Qy    1627  attgttccgcctcacgtttgaacgtgaaccgagccagacataccttccac  1686
Db   37902  ATCGTCCCGGCACGCGTTGCATCGCCCAAACTCGGGTTCCCCAATCCGGCAT  37843

Qy    1687  tggcttgccggcgatgagctgttcggttcggggcgcaggagacaccattaggaatccggtact  1746
Db   37842  TGGCTGCGCGGCCGCGCGGCAACTGCTGGAAGTGGCCGTACCGCGCTGGTGAACTCGTCGCAGGCC  37783

Qy    1747  gaagatatcttcaacaagcaggctgtgctggatatgctgaacgagcaccgcgatggcgtg  1806
Db   37782  GAACATCTGGTCAACCTTAGCCCGTGCGATGCTTGACGAGCACCGAGGTGGCATC  37723

Qy    1807  tcagatcattcccgtcgactggactgttcgtcattatgtgtggcacggcattttt  1866
```

```
          ||||||  || || ||||||||| || ||  |||||| |||||||||| ||| ||
Db 37722 AGCGATCATAGTCGCCGGCTGTGGACGGTGCTAATTTTATGCTGTGCACGCCATCTTC 37663
          |||||| || ||||| | ||||||||      || |||||| |||
Qy  1867 gtggaaaccgcattgatccacagattgaggaccggctcctacccggtcgagc 1918
          |||| || | |  |||||| |  || |||| || | ||||| |||
Db 37662 ATCGAACACAGTCTGGTGCCACAGATCAGCGAGCCACAGTATCCGGTACAGC 37611

RESULT 14
LOCUS       AF168003               10772 bp    DNA     linear   BCT 24-JUL-2000
DEFINITION  Desulfovibrio gigas putative PhoP (phoP), putative response
            regulator (phoR), putative Soj (soj), hypothetical protein,
            flavodoxin (fld), hypothetical protein, and putative asparagine
            synthetase (asn) genes, complete cds.
ACCESSION   AF168003
VERSION     AF168003.1  GI:6978027
KEYWORDS    .
SOURCE      Desulfovibrio gigas.
  ORGANISM  Desulfovibrio gigas
            Bacteria; Proteobacteria; delta subdivision; Desulfovibrio.
REFERENCE   1  (bases 1 to 10772)
  AUTHORS   Broco,M., Oliveira,S., Silva,G., Agostinho,M. and
            Rodrigues-Pousada,C.
  TITLE     A DNA fragment of Desulfovibrio gigas genome containing replication
            origin related genes
  JOURNAL   DNA Seq. 11 (1-2), 119-124 (2000)
  MEDLINE   20358724
   PUBMED   10902918
REFERENCE   2  (bases 1 to 10772)
  AUTHORS   Broco,M., Oliveira,S., Silva,G., Agostinho,M., LeGall,J., Xavier,A.
            and Rodrigues-Pousada,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUL-1999) Genetica Molecular, Instituto de Tecnologia
            Quimica e Biologica, Rua da Quinta Grande, 6, Oeiras 2780-156,
            Portugal
FEATURES             Location/Qualifiers
     source          1. .10772
                     /organism="Desulfovibrio gigas"
                     /strain="ATCC19364"
                     /db_xref="ATCC:19364"
                     /db_xref="taxon:879"
     gene            1400. .1786
                     /gene="phoP"
     CDS             1400. .1786
                     /gene="phoP"
                     /codon_start=1
                     /transl_table=11
                     /product="putative PhoP"
                     /protein_id="AAF34246.1"
                     /db_xref="GI:6978028"
                     /translation="MGKKLLIIEDELHIRTLLEQALEDLADEFAVEILTAADGEEGLA
                     LIKTERPQVVFLDIMMPKMNGYEVCTAVKNDPELKDTVVVLLTAKGQEADKKKGLEIG
                     AYDYMTKPFDPDEVVDLAKELLQIVD"
     gene            2238. .4328
                     /gene="phoR"
     CDS             2238. .4328
                     /gene="phoR"
                     /codon_start=1
                     /transl_table=11
                     /product="putative response regulator"
                     /protein_id="AAF34247.1"
                     /db_xref="GI:6978029"
                     /translation="MLERDAIAAFPGFFPQARLICSARVGFSSFMLFLEIPPDIVALA
                     NAAHMFRMQSRERLAMKLPSQVAAACDIHVGYAYVDSQPAGMDSVLLRAYLHAQRLG
                     ARPRDVASPLHKEFQCIVERGKVTTLYQPIVDFETSQVMGWEAFSRGPENFFSSPQ
                     MLFQLAREAGATFRLEKLCREKALTGLGQVGSNQRLFLNNHSDEIADPNFTPGLLEE
                     IEAYGLKPHNIVLEFAERQGFKDMNLFFRDIERFRAKGFAIAIDDVGSGDSSLRNITL
                     IRPDFIKFDISLVEGIDTNPYHRSMVETLLFLAQRIGARTIAVGIETETALSLSVSMG
                     VNAGQGNCLGVGPGYKPASPCVAIPCKTGFTQADGQSWKCSAPIRELVMPCPTTSENAT
                     VREIKEQLADKPAQTSMVVVRNGRPCGLLMQYNTDRHLSTQFGMSLYYNREVARIMDP
                     APLIVEQDLPIEDVAALAMVRESTKVYDDIIVDVEKLLGVVSVQRMLDLQTQVEL
                     AKGANPLSGLPGNVSIEREIERRLKGCHPTAMAYIIDLDNFKVYNDVVGFERGDRIILL
```

```
                     TADLLREAVLCGGEGDFLGHIGGDDFLVMADPARVDICNHIMHHFALRSKELYSPE
                     DAARGYLVGKGRDGQTGRFPLTSLSIAVIDCPMDACFSMELLSNRTAAVKHAKSVEG
                     NSLVREPLLAATAA"
     gene            5063. .5794
                     /gene="soj"
     CDS             5063. .5794
                     /gene="soj"
                     /note="partitioning protein"
                     /codon_start=1
                     /transl_table=11
                     /product="putative Soj"
                     /protein_id="AAF34248.1"
                     /db_xref="GI:6978030"
                     /translation="MISKGGVGKTTAVHLAAGLALNNYKTLLVDCDTQGQAAFMLGI
                     RPNDGLADLILGDTPPDKAMLKARDNLWLLSGGKSLAKIKRHITRKDYAGENTLAEAL
                     RPIERAFQFFILDTSPSWDSLTVNVLFYAQELLAPISLEVMSIQGLAEFLRNFAALRK
                     HNPGVTLKYLLPMHLSKPSKNSQAILESLEKFYGKYLCTPIRYSPKLAEAPAYGMTVY
                     EYAGSDNVVKDFRELLKDVLHKDEL"
     CDS             5851. .7206
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="hypothetical protein"
                     /protein_id="AAF34249.1"
                     /db_xref="GI:6978031"
                     /translation="MYESLTISMGPQAPALRIPRAKVEDRHAGGAPRLLRLYDPTPLE
                     TPLGRLIPRHTDDSGEPENRRPARPAIEILTPAGTVKALSLEEQTLVTTLGPQPAERL
                     LFYPDGALRRLRFPQDGKLNAYLSWQDEKATAPLMRLELPPDPNVLEAGRRVLQARCIS
                     LLFYPSGALQSVSLWTGEVVPWPTPVGTLPARIGLAFHENGALASLEPARPVEILTPL
                     GPVTVFDPDPEGISGDINSLAFCPQGRLERLTTASHLIRSTTDPAKSISPSRQLSMCS
                     EDEEWISMPLTVCFRGDAVLAGLGSLGASGVLPPTLQLSLQLDQPEAAPAGLPQPH
                     GLRQPGGLRLSGAAAEMPLGGPRSFCRNSSPQPPSSKTFVIITKYYQTFGRERERG
                     EPFVKRFPLSQRPFSKLKCPIPAPALPWFSPAPRQSPARRRSSASLPPGLRPIPRPSS
                     R"
     gene            complement(7085. .7525)
                     /gene="fld"
     CDS             complement(7085. .7525)
                     /gene="fld"
                     /note="similar to Desulfovibrio gigas flavodoxin encoded
                     by GenBank Accession Number X64766"
                     /codon_start=1
                     /transl_table=11
                     /product="flavodoxin"
                     /protein_id="AAF34250.1"
                     /db_xref="GI:6978032"
                     /translation="MPKALIVYGSTTGNTEGVAEAIAKTLNSEGMETTVVNVADVTAP
                     GLAEGYDVVLLGCSTWGDDEIELQEDFVPLYEDLDRAGLKDKKVGFGCGDSSYTYFC
                     GAVDVIEKKAEELGATLVASSLKIDGEPDSAEVLDWAREVLARV"
     CDS             complement(7675. .8466)
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="hypothetical protein"
                     /protein_id="AAF34251.1"
                     /db_xref="GI:6978033"
                     /translation="MTAPTDNAAPGAPRRDFLDLDIPLEGVTAYWLSLHKLLEPHKKL
                     KNAQRQIQDEAQYTTEPLVRYLLEVLVSDMDAPDMRRLGQARADMLLDGVARRLQLMR
                     VALMDMATAENPRRTLAKMTAQHSMPLIDEKAFSLAQQLLGLAATPGWEAKVHYFAV
                     DHRIKDEKLLVALLFHAANARREGKMAARQFLPHLHSVFFQQGLALVIDGFDAPFLRK
                     RLKVHQAAIMADARRKTACSLELACAIKQRLPYDTAFTMAKAWLA"
     gene            complement(8687. .10618)
                     /gene="asn"
     CDS             complement(8687. .10618)
                     /gene="asn"
                     /codon_start=1
                     /transl_table=11
                     /product="putative asparagine synthetase"
                     /protein_id="AAF34252.1"
                     /db_xref="GI:6978034"
                     /translation="MCGIAGFVLLQDAPPVTDHQARDVLAAMAGRIAHRGPDGRGHHV
                     GLPAAFAHCRLAIIDLSTGAQPMAAEDGAAVVTFNGEINYQEVRRELQALGRRTFQTR
                     SDTEVLLQAYLEWGVACLDRLEGMFAFALWDVRRRTLFAARDRFGSKKPFFYTIQQGVF
                     AFASELSALEAVPWAAFSIRASTLARFLAYEYVPTPESIYAEARKLPPAQVLEHEDG
```

```
          TLLVERYWETPMPRPEGSKGFGRADEPALCEELRRLLTQAVRRRLIADVPVGVFLSGG
          VDSSSVAAMAGQLAPGIKTFSIGFTEASYDESAYAGRVAAQLGTDHQLELLSADACAD
          LLPEVVSRFDEPMADPSIVPTYLLSQITTRRQVVTVALGGDGPDELLAGYEYYPGFKATQ
          YWERLPAWLRRAAVAASGLLPGSAGYVNPRHVAQTFLAAAEAPAWLRTQRMLTALTPE
          AQRALWKSPGDMLEPTRLFASTKRLWEALLPDWPPMARVFHLFAQITYGYILVKVDRC
          TMLHGLEARAPFLDRDLAEFCFRLPVSMKLRGATRKYLLKEAMAGLLPQDILHRPKRG
          FLIPTAAWLRGRLRPLVDDLLHERRLREQGLFHPAAVRRLVEEHASGKKDHRKGLWTM
          LVLQIWLAGKPAVRIDPER"

BASE COUNT    2074 a    3596 c    3223 g    1879 t
ORIGIN

Query Match            9.2%;    Score 177;    DB 1;    Length 10772;
Best Local Similarity    50.9%;    Pred. No. 1.3e-31;
Matches 547;    Conservative 0;    Mismatches 495;    Indels 33;    Gaps 4;

Qy     80  gcatgcgcaccggtggtcctgacgatgccggcacttggcatgacgccgatgcagccgtttg   139
           |||||| ||  ||||||| |||||||||||| | ||  |||| |   | ||  ||  ||||||
Db  10527  GCATCGCCCCACCGGGTCCCGACGGCCGCGGCCATCATGTGGGCCTGCCCGCCGGCCCTTTG   10468

Qy    140  gattcaaccgcctctccatcattgcacactcccaaccactcgtttggggac----------   199
           ||| ||  |||||||   |||||||||||||||||| |||||||||||||||
Db  10467  CCCATTGCCGGCTGGCCATCATGCAGCCTCTCCACCGCGGCCAGCCCATG----------   10416

Qy    200  ctgcggatgaaccgaccgctacgcaatgacttcaacgtgagatctacaactacgttg   259
           | |||  ||||||||  |||||||| ||||||||||||||||||||||||||||||
Db  10417  --GCCCGCGAGGACGCGCGGCCGTGTGACCTTCAACGGCGAGATCTACAACTATCAGG   10360

Qy    260  agctcgctaaagagctctcggattgggatatacctttaatacttctggcgatggcgagc   319
           ||||||| |||| ||||||||| ||||||| | ||||||||||| ||||||||||||||
Db  10359  AAGTTCGCCGCAGCCGAGCTACAGGCCCTGGGCCGGCGACCTTCCAGCACCCGTTCGGATACGGAAG   10300

Qy    320  caattgttgtcggttccaccactggggcgagtccgttccgtggtcgagcatctccggaatgt   379
           |||  | |||| ||||| | ||||||||| ||||||||||||||||||||||||||||||
Db  10299  TCCTGTTGCAGGCCTATCTGGAATGGGAGCTGCTGACCGGCTGGAGGGCCATGT   10240

Qy    380  tcggcattgccatttgggatacaaaggaaaagtcgctttcctgcgcgtgatcagttcg   439
           ||| ||||||||||||||||| || ||||||||||||||| ||  |||||||| ||||
Db  10239  TCGCCTTGCCCGTGTGGATGTCCCGCCGACCGCTCTTTCCGCCGGGGACCCGCTTTG   10180

Qy    440  gcatcaagccactgttctacgcaaccaccgagcatggcacccgtgttctcctcagagaga   499
           |||| || ||||||||||||| || ||||||  |||  ||||| |||||||||||||||
Db  10179  GCAAGAAGCCGCTTTTTACGATCCAGCAGGGCGTGTTCGCCTTCGCCCTCGGAGCTTT   10120

Qy    500  agaccatcttggagatggcgcgaggagatgaatctagatctccggcgccttgataagcgcacca   559
           |||||| ||| || ||||||| |||||  | || ||||||||   || |||||||||
Db  10119  CGGGCCCTGGAGCGGCCGCTGCGCGGCCGT-TCTCCATCCGGCCCTC-------CACCC   10069

Qy    560  ttgagcactacgtggacctgcagctgcgccgagctgtgccccgagaccttcacgcgcagattt   619
           | ||| ||||||| || |||||||| |||||||| |||  ||||||| |||||| ||  ||
Db  10068  TGGCCGCTTTTTGGCCTACGAGTATGTGCCCACGGCGAAAGCATCTACCGCCGAGGCGC   10009

Qy    620  ccgcgcttgggagtcaggctgcgaccgcaacagttcgtccgggcggcaacctgaacga   679
           | ||||| ||||| |||||||||||| |||| ||||||||||||||||||||||||||
Db  10008  GCAAACTGCCGCGGGTGGAGCTGCAGCTGGCGCTGTACTGGAATCGGCCCACGCGTGTTGGAAC   9949

Qy    680  gttacttcaagcctcagttcccagtacagaaggtcgtaaagggt----------aagg   727
           |||||||||| ||| |||| |||  |  | ||||||||| ||||
Db   9948  GCTACTTCGAGGAGAGCCGCCAGTATCCCGAGGGGGTGGCAGGCAAGGCTTTGGCCGGGGGCACG   9889

Qy    728  agcaggacctcttcgatcgacattgcacggcattgcccagtgtggaggatagcgtcgaaaagcatatgc   787
           ||| || |||| |||| |||| |||||  || || ||||| ||||||||||| ||||| || ||  ||
Db   9888  AGCAGGAGCCCCTGTCGCGAGGAGTTGCCGGCGTTGCGCTGACCCAGGCCGCTGCACCAGGCCGCCTCA   9829

Qy    788  gtgcgacgtagctaggctcgttccttccgggcattgactcaaccgcaattgcgc   847
           |||||||||| ||||| |||||| | ||||||| |||||||  ||| ||||| |
Db   9828  TCGCCGATGTGCCCGTGGCGCAGTGCGGCGGTGTTCCTGTTCCCGTGCAACGGTCCTCCGTGGCGG   9769

Qy    848  cgcttgcaaagcggccacaaccgtcgacctgctcacctcaccaccggtttcgagcgtgaag   907
           ||  ||||||||||||| ||||||||| |||||| | |||| || ||||||||| || |
Db   9768  CCAATGGCGCCAGCTGCGGCCGGGCATCAAGACCTTTTCCATCGGCTTCACCGAGGCCA   9709

Qy    908  gctactcggaggtcgatgtggctcggaggctccggaggctggccgttcggcggcacatcg   967
```

```
Db   9708  GCTACGACGAATCGCCTATGCCGCGGCGCCGGGTGGCCGCGGCCAGTGGGACGCGGATCACCAGC   9649

Qy    968  tgaagattgtctcgcctgaggaatacgccaacgcgattcctaagatcatgtggtacttgg   1027
           |||| || |||||| |||||||| | |||||||||||| ||||||| |   |  ||||||
Db   9648  TGGAACTGCTTTCGGCCGGATGCCGCCGACCTGCTGCCCGAGGTGTGTCCCGGTTCG   9589

Qy   1028  atgatcctgtagctgaccatcattggtcccgctacttcgtggcagcggaagcacgta   1087
           ||||| |||   | || |||||   || |||||   |||||||||| |  ||  ||||
Db   9588  ACGAGCCCATGGCCAGCCGACCCCTCCATCGTGCCCATGCTGCGTCCCAGATCACCCGGC   9529

Qy   1088  agcacgtcaaggtgtgtctcgtcgtgcgcgagggcgcagagagctgttcggtggata   1142
           ||||||| |||| || | | |||| |||||||||  ||||||||||||||  ||||||
Db   9528  GACAGGTCACCGTGGCCCTGGCCGGCGCGATGCGGCCCGATGAACTGTTGCCGGGTA   9474
```

```
RESULT 15
AB039379

LOCUS       AB039379               5188 bp    DNA    linear    BCT 05-SEP-2000
DEFINITION  Streptomyces aureofaciens tcsD, tcsE, tcsF, tcsG, genes for
            beta-ketoacyl synthase, beta-ketoacyl synthase2, acyl carrier
            protein, asparagine synthase homolog, complete cds.
ACCESSION   AB039379
VERSION     AB039379.1    GI:9971571
KEYWORDS    asparagine synthase homolog; acyl carrier protein; beta-ketoacyl
            synthase2; beta-ketoacyl synthase.
SOURCE      Streptomyces aureofaciens
  ORGANISM  Streptomyces aureofaciens (strain:NRRL3203) DNA.
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (sites)
  AUTHORS   Nakano,T., Ikeda,M., Mizukami,T. and Katsumata,R.
  TITLE     Analysis of the chlortetracycline polyketide synthase genes from
            Streptomyces aureofaciens NRRL3203
  JOURNAL   Unpublished (2000)
REFERENCE   2  (bases 1 to 5188)
  AUTHORS   Nakano,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-FEB-2000) Tetsuo Nakano, Kyowa Hakko Kogyo, Co.,
            Ltd., Technical Research Laboratories; 1-1 Kyowa-cho, Hofu,
            Yamaguchi 747-8522, Japan (E-mail:t.nakano@kyowa.co.jp,
            Tel:81-835-22-2518, Fax:81-835-22-2466)
FEATURES             Location/Qualifiers
     source          1..5188
                     /organism="Streptomyces aureofaciens"
                     /strain="NRRL3203"
                     /db_xref="taxon:1894"
     gene            443..1726
                     /gene="tcsD"
     CDS             443..1726
                     /gene="tcsD"
                     /codon_start=1
                     /transl_table=11
                     /product="beta-ketoacyl synthase"
                     /protein_id="BAB12566.1"
                     /db_xref="GI:9971572"
                     /translation="MVGVDDGRRRVVVTGIGVVAPGEVGTKAFWELLTAGRTATRTIS
                     TFDPSPFRSRVAAECFDPQAAGLTQRQIRQWDRTAQFAVVAAREALDDSGVSGVNDP
                     IRTGVMAGTACGMTISLDREFAVVSDEGRKRLVDDAHGVPYLYDYFVPSSMATELAWL
                     VEAEGPVGVVSGTCSGIDVIAHACDLIRDDAADVMLAGASDAAISPITVACFDAIKA
                     TTPRNDEPESASRPFDLTRNGFALGEGAAMFVLEEREHAIRRGAHIYAEVTGYAGRCN
                     SYSMTGLRVDGNEMAEAIRVALDRSRMNPEDVGYVNAHGSATKQNDRHETAAFKKSLG
                     DHAYRVPVSSIKSMIGHSLGAVCALEFAASALAIEHSVVPPTANLHVPDPECDLDYVP
                     LVAREQRVRTVLSVASGFGGFQSAIVLAEPEWSRG"
     gene            1723..2928
                     /gene="tcsE"
     CDS             1723..2928
                     /gene="tcsE"
                     /codon_start=1
                     /transl_table=11
                     /product="beta-ketoacyl synthase2"
                     /protein_id="BAB12567.1"
                     /db_xref="GI:9971573"
```

```
                  /translation="MTDVVTGIGVVAPNGWGVAEYWDATLAGRSGLGPLTRPDSTGY
                  RAHVAGEVNFTPEEHLPGRLLPQTDHMTRLALVAAEEALADAGADPTTMPDFSAGVVT
                  AASAGGFEFGQKELQALMSKGPAYVSAYQSFAWFYPVNTGQISIRHGMRGPGSALVAE
                  QAGGLDALAKARRHVREGTALMLAGAVDGSLCPWSWLCLSRSGTVSPRRDPDQRAFLPF
                  DADASGGVPGEGALLVLETPDGAAERGQHRHYGRIAGYAATFDPRPGSGRPPGLRRA
                  IDTALADAGIGPSDVDVVFADASGVAEQDRAEAEVLGAVFGPRGVPVTAPKTMTGRLL
                  AGGAPLDVAAALLSLRDGVIPPSVHIGRPAHPDLDLVRDRPRPARLGSALVLARGHGG
                  FNSALVLTS"
gene              2998..3252
                  /gene="tcsF"
CDS               2998..3252
                  /gene="tcsF"
                  /codon_start=1
                  /transl_table=11
                  /product="acyl carrier protein"
                  /protein_id="BAB12568.1"
                  /db_xref="GI:9971574"
                  /translation="MAEFTLDDLIAALKEAAGADEGVAHDDILDVPFADLYDSLALF
                  NTLNTIERERGITLPDETVVEATTPRMLLDAVNLGLATAG"
gene              3255..5084
                  /gene="tcsG"
CDS               3255..5084
                  /gene="tcsG"
                  /codon_start=1
                  /transl_table=11
                  /product="asparagine synthase homolog"
                  /protein_id="BAB12569.1"
                  /db_xref="GI:9971575"
                  /translation="MCGIVGVVDYDAPAEHRRDLLQAMTDTMACRGPDAEGLWFGRAA
                  GLGHRRLSVIDPEHGRQPMTAEHEGRGAAITFSGEIYNFRELRAELTSHGHRFRTSCD
                  TEVVLRGYLQWGAALVERLNGMFAFAVWDEHSEELLLVRDRMGVKPLYYWPTAAGVRF
                  GSEPKAVLADRTLSRRVGPDGLCEVLDMVKTPEAAVFSGLVEVRPGQLVRVGRGGVRR
                  QTYWRLEAREHTDDLPTTIGTVRDLLADTVTRQLVSDVPLGTLLSGGLDSSAVTALAA
                  RSGPVRSFSVDFRGAADAFASDAVRREADAPYVRELAGHVGADHLEVLLDSPGLSDPA
                  VRSAVLRATDLPPAYWGDMWPSLYLFFRAVRENVTVALSGEAADELFGGYRWFRNPLA
                  LRAATFPWLTPGSARYFGGSSLFQAGLLDKLDLDGYRAARYAEALAEVPVLPGESAPE
                  RRMREIGHLNLTRFVQTLLDRKDRMSMAVGLEVRVPFCDHRLVEYVFNVPWEMKSFDG
                  REKSLLRAAVADLLPRSVVERVKTPYPATQDPGYEGALRAELAGVLADPDAPVLPLLD
                  RERALAAAARPTGAVSRPYDRGSLELVLWLNSWLAEYEVTLDL"
BASE COUNT      677 a    1873 c    1944 g     694 t
ORIGIN
```

```
Query Match             6.1%;   Score 116.6; DB 1;  Length 5188;
Best Local Similarity  56.1%;   Pred. No. 3.4e-17;
      Conservative  243;   Conservative   0; Mismatches  184;  Indels    6;  Gaps      1;

Qy      81 catcgcgcaccggtgtcctgacgatgccggcacttggcacttggcggcatgacggcagcgtttgg 140
           ||-|||-|||-|| | |||||||| |-|||| |-||||-|| |--||||||||||||||||||
Db    3335 CATGGCCTGCCGGGGGGCCCGACCGCCAGGGCCTCTGGTTCGGCCCCGCCGGCGGCCGGCTCGG 3394

Qy     141 attcaaccgcctccatcattgatattgcacactcccacaacactcgttggggacc 200
           |-|||||| | |||||| | |||||||-|| |-|||-|| |-||| ||||||| |-||||
Db    3395 TCACCGGCCCTGCGGTGATCGACCCCGAGCACGGCCGCCGATGACCCGGAAC- 3453

Qy     201 tgcggatgaaccgaccgctacgcaatgacttcaaacgtgagatctacaactacgttga 260
           |-|| ||| | || |-|||| |||||-||||||-|| |||||||||||||||||||
Db    3454 -----ACGAGGGCCCGGCGGCGGGCGGCGGATCACCTTCAGCGAGCGGAGATCTACAACTTCCGTGA 3508

Qy     261 gctgcgtaaagagctctcggatttgggatataccttaatacttctggcgatggcgagcc 320
           ||||-|| |-|||| | |||| |-|| |-| | |-||-||||||| ||
Db    3509 GCTGCGGCCGAACTCACCTCGCACAGGCCACCGTTCAGGACGTCCTGCGATACCGAGGT 3568

Qy     321 aattgttgcggttccacactggggcgagtccgtcggtcgagcatctccgcggaatgtt 380
           ||| |||-|| | |||||||-||| |||-|| ||-|| |||||| |||-|||| |||||
Db    3569 CGTCTCGCGCGGCTACCTCCAGTGGGGCGCCGGCGCGCTCGAGAGGCTCAACGGGATGTT 3628

Qy     381 cggcattgccatttgggatacaaaggaaaagtcgctttccttgccggtcgatcagttcgg 440
           ||-| | |||||| |||||-|||-||||-| ||||| |||-|| ||||||| |-| ||||
Db    3629 CGGCTTCGCCGTCTGCTGGGACGAGCACAGCAGGAACTCCTGCTCGTCCGCGACCGGATGGG 3688

Qy     441 catcaagccactgttctacgcaaccaccgagcatgcacgcgttctcctcagaagagaa 500
           | | ||||-|| |-|||-||-|| |||-|||| |-||||||| ||| ||| || ||
Db                                                                500
```

```
Db    3689  CGTCAAGCCGCTCTACTACTGGCCGACCGCCGCCGGCGTGCGCTTCGGCTCGGAGCCCAA 3748

Qy     501  gaccatcttggag 513
            | ||-||| ||| |
Db    3749  GGCCGTCCTGCGG 3761
```

Search completed: September 20, 2002, 07:46:43
Job time: 8395 sec

```
                    GenCore version 4.5
           Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    September 20, 2002, 05:55:13 ; Search time 234.38 Seconds
                                         (without alignments)
                                         14064.663 Million cell updates/sec

Title:          US-09-786-474-1
Perfect score:  1920
Sequence:       1 atgtgcggcctttcttggcat.......gctcctaccggtcgagctt 1920

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

arched:         1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters:    3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      N_Geneseq_032802:*
      1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
      2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
      3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
      4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
      5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
      6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
      7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
      8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
      9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
     10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
     11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
     12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
     13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
     14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
     15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
     16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
     17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
     18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
     19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
     20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
     21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
     22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000A.DAT:*
     23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
     24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1920	100.0	1920	21	AAA10292	DNA encoding Coryn
2	1920	100.0	3825	21	AAA10293	DNA encoding Coryn
3	1904	99.2	1920	22	AAH67371	C glutamicum codin
4	1904	99.2	349980	22	AAH68531	C glutamicum codin
5	1902.4	99.1	1962	22	AAF71807	Corynebacterium gl
c 6	116.6	6.1	30001	18	AAT61016	Total DNA sequence
c 7	116.6	6.1	30001	20	AAX05110	S. aureofaciens DN
c 8	77.4	4.0	1872	22	AAF77894	Quorum sensing con
c 9	74.6	3.9	349980	22	AAH41225	Pyrococcus abyssi
10	67.8	3.5	58857	21	AAA58471	Nucleotide sequenc
11	45.6	2.4	1470	21	AAC78035	Human cancer assoc
12	45.6	2.4	1518	23	AAS64941	DNA encoding novel
13	45.6	2.4	3331	22	AAS22442	Human cDNA encodin
14	45.6	2.4	3690	22	AAS22678	Human cDNA encodin
15	44.4	2.3	1836	21	AAC46946	Arabidopsis thalia
16	42.8	2.2	2162	21	AAC50959	Arabidopsis thalia
17	41.6	2.2	2164	21	AAC40911	Arabidopsis thalia
18	41.2	2.1	1664976	19	AAV21209	Methanococcus jann
19	37.6	2.0	3837	23	ABL04631	Drosophila melanog
c 20	37.6	2.0	1349	21	AAZ51123	Human purino recep
c 21	37.6	2.0	1349	22	AAH25209	Nucleotide sequenc
c 22	37.6	2.0	1421	21	AAZ51322	Human purino recep
c 23	37.6	2.0	1421	22	AAH25208	Nucleotide sequenc
c 24	37.6	2.0	1436	21	AAZ51321	Human purino recep
c 25	37.6	2.0	1436	22	AAH25207	Nucleotide sequenc
c 26	37.6	2.0	1499	21	AAZ51324	Human purino recep
c 27	37.6	2.0	1499	22	AAH25210	Nucleotide sequenc
28	36.8	1.9	1323	23	AAF55439	Nucleotide sequenc
29	36.8	1.9	1872	23	AAS52741	E. coli DNA for ce
c 30	36.8	1.9	2047	23	AAS79318	DNA encoding novel
31	36.8	1.9	3192	23	AAS93937	DNA encoding novel
c 32	36.8	1.9	3192	23	AAS80208	DNA encoding novel
c 33	36.8	1.9	3192	23	AAS89446	DNA encoding novel
34	36.8	1.9	19717	23	AAS93358	DNA encoding novel
c 35	36.8	1.9	242	23	AAS59938	Propionibacterium
c 36	36.6	1.9	1542	22	AAH82183	Rat differential t
37	36.4	1.9	1542	21	AAZ60717	DNA encoding a bet
38	36.4	1.9	2557	22	AAD14504	S. clavuligerus cl
39	36.4	1.9	11604	15	AAQ62625	Rat N-acetyl-gluco
40	36.4	1.9	15079	16	AAD14501	Streptomyces clavu
41	36.4	1.9	15079	22	AAQ91580	S. clavuligerus cl
42	36.2	1.9	1239	21	AAD14499	Streptomyces clavu
43	36.2	1.9	1242	21	AAC45664	Arabidopsis thalia
44	36.2	1.9	1313	22	AAC33977	Arabidopsis thalia
c 45	36.2	1.9		22	AAF74867	Leishmania major P

ALIGNMENTS

```
RESULT  1
AAA10292
ID   AAA10292 standard; DNA; 1920 BP.
XX
AC   AAA10292;
XX
DT   03-JUL-2000 (first entry)
XX
DE   DNA encoding Corynebacterium glutamicum lysozyme insensitivity protein.
XX
KW   Lysozyme insensitivity protein; bacterium; recombinant expression;
KW   amino acid production; ds.
XX
OS   Corynebacterium glutamicum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1920
FT                   /*tag= a
FT                   /product= "Corynebacterium glutamicum lysozyme
FT                   insensitivity protein"
FT                   /note= "No stop codon given in the specification"
XX
PN   WO200014241-A1.
XX
PD   16-MAR-2000.
XX
PF   04-SEP-1998; 98WO-JP03981.
XX
PR   04-SEP-1998; 98WO-JP03981.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
```

PI   Nagai K, Wati M;
XX
DR   WPI; 2000-256989/22.
DR   P-PSDB; AAY87459.
XX
PT   New DNA molecule useful for production of amino acids encodes protein
PT   imparting lysozyme resistance to Corynebacterium glutamicum strain –
XX
PS   Claim 3; Page 33-35; 50pp; Japanese.
XX
CC   This sequence represents the DNA coding sequence encoding a
CC   Corynebacterium glutamicum lysozyme insensitivity protein. This protein
CC   renders lysozyme-sensitive strains of Corynebacterium glutamicum
CC   insensitive to lysozyme. DNA sequences encoding the lysozyme
CC   insensitivity protein can be used to construct a recombinant vector for
CC   the expression of this protein in a host cell. The DNA encoding the
CC   lysozyme insensitivity protein can also be mutated, and used to generate
CC   strains of Corynebacterium glutamicum in which this protein is
CC   inactivated and which are therefore lysozyme-sensitive. Corynebacterium
CC   glutamicum is used for the production of amino acids, in particular
CC   glutamic acid and glutamine.

Sequence 1920 BP; 425 A; 535 C; 524 G; 436 T; 0 other;

Query Match   100.0%;   Score 1920;   DB 21;   Length 1920;
Best Local Similarity 100.0%;   Pred. No. 0;
Matches 1920;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

```
Qy         1 atgtgcggccttcttgactgcatattgactgacggcaaatggaacgctgaagcattcctgca 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 atgtgcggccttcttgactgcatattgactgacggcaaatggaacgctgaagcattcctgca 60

Qy        61 ctcgagcgggccttgccatgcatgcgccaccgtgctcctaacgatgccggcacttggcat 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 ctcgagcgggccttgccatgcatgcgccaccgtgctcctaacgatgccggcacttggcat 120

Qy       121 gacgccgatgcagcgtttgagctgcgtaaagagctctcggattcacactccac 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 gacgccgatgcagcgtttgagctgcgtaaagagctctcggattcacactccac 180

Qy       181 caaccactgcgctggggacctgcgcggatgaaccgacgctactgcaatgacttcaacggt 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 caaccactgcgctggggacctgcgcggatgaaccgacgctactgcaatgacttcaacggt 240

Qy       241 gagatccacaactacgttgagctgcgtaaagagctctcggatttgggatatacctttaat 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 gagatccacaactacgttgagctgcgtaaagagctctcggatttgggatatacctttaat 300

Qy       301 acttctggcgatggcgagcaattgtgtcggttccaccactggagttccggttc 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 acttctggcgatggcgagcaattgtgtcggttccaccactggagttccggttc 360

Qy       361 gagcatcccggaatgttcgcgcattgccatttggatacaaagaaagtcgctttc 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 gagcatcccggaatgttcgcgcattgccatttggatacaaagaaagtcgctttc 420

Qy       421 cttgcgcgtgatcagttccggtccgatctgttctacgcacagccatgcacc 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 cttgcgcgtgatcagttccggtccgatctgttctacgcacagccatgcacc 480

Qy       481 gtgttctctcagaagaagaagaccatcttgagatggccgaggagtggcacctg 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 gtgttctctcagaagaagaagaccatcttgagatggccgaggagtggcacctg 540

Qy       541 ggcctgataagcgcaccttgagcactacgtgccctcagctgccgagccagat 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 ggcctgataagcgcaccttgagcactacgtgccctcagctgccgagccagat 600

Qy       601 acccttcacgcgcagatttcccgctcagctggatcaggctgacagttcgtccgggc 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 acccttcacgcgcagatttcccgctcagctggatcaggctgacagttcgtccgggc 660

Qy       661 ggcaagctggaacagagaagcgttacttcaagcctccagtcccagttcccagtacagaag 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 ggcaagctggaacagagaagcgttacttcaagcctccagtcccagttcccagtacagaag 720

Qy       721 gtaaggagcaggacctccttgcatcgcattgcgatggatagcgtgagttagcgtcggaaag 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 gtaaggagcaggacctccttgcatcgcattgcgatggatagcgtgagttagcgtcggaaag 780

Qy       781 catatgctgccgacgtgaccgtaggctgcgtccttccggcggcattgactcaaccgca 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 catatgctgccgacgtgaccgtaggctgcgtccttccggcggcattgactcaaccgca 840

Qy       841 attgcgccgcttgcaaagcgccaacacctgactcgtcacccttcaccacacggtttcgag 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 attgcgccgcttgcaaagcgccaacacctgactcgtcacccttcaccacacggtttcgag 900

Qy       901 cgtgaaggctactcggagtcgatgtgctggctggctgcgtcggagtccgcggtggcgctgag 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 cgtgaaggctactcggagtcgatgtgctggctgcgtcggagtccgcggtggcgctgag 960

Qy       961 cacatcgtgaagattgtctgcgctgaggaatacgccaacgcgatcctaagatcatgtgg 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 cacatcgtgaagattgtctgcgctgaggaatacgccaacgcgatcctaagatcatgtgg 1020

Qy      1021 tacttggatgatcctgagctgaccgacccatcatggtcccgctgtacttcgtgcagcggaa 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 tacttggatgatcctgagctgaccgaccccatcatggtcccgctgtacttcgtgcagcggaa 1080

Qy      1081 gcacgtaagcacgtcaaggttgctgctcggcagggcgcagatgagcgtgtcggtgga 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 gcacgtaagcacgtcaaggttgctgctcggcagggcgcagatgagcgtgtcggtgga 1140

Qy      1141 tacaccatttacaaagacgcgcatctcgctgcctacgatcgacacttggcat 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 tacaccatttacaaagacgcgcatctcgctgcctacgatctccctccccacta 1200

Qy      1201 cgtaaaggctggaaagctcagcaggttctgccagacggcgatgagggcaagtcccttt 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 cgtaaaggctggaaagctcagcaggttctgccagacggcgatgagggcaagtcccttt 1260

Qy      1261 cttgagcgtggctccatgacctacgcaatgagagctactacggcaacgctcctcaat 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1261 cttgagcgtggctccatgacctacgcaatgagagctactacggcaacgctcctcaat 1320

Qy      1321 ttcgagcagatgcaacgcgttattccatggcaaagcgcgaatggacaccgcgaagtc 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1321 ttcgagcagatgcaacgcgttattccatggcaaagcgcgaatggacaccgcgaagtc 1380

Qy      1381 actgcaccgatctacgcacgacaatccgcaacttgatccgtagtagcccgatgcaacacctg 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1381 actgcaccgatctacgcacgacaatccgcaacttgatccgtagtagcccgatgcaacacctg 1440

Qy      1441 gatctgttcacctggatgcgcggcgacatccgtcaaggctgacaagatcaacatggcg 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 gatctgttcacctggatgcgcggcgacatccgtcaaggctgacaagatcaacatggcg 1500

Qy      1501 aactcccttgagctgcgagttccattctggataagtgagaagtttcaagttgagagacc 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1501 aactcccttgagctgcgagttccattctggataagtgagaagtttcaagttgagagacc 1560

Qy      1561 attcctttacgatctgaagattgccacacaagtaccaccgtaccacctgcgctgcgcaggggcactc 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1561 attcctttacgatctgaagattgccacacaagtaccaccgtaccacctgcgctgcgcaggggcactc 1620

Qy      1621 gagcagattgttccgcctcacgtttgcaccgcaagaagctggcttcccatggcttcccatg 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1621 gagcagattgttccgcctcacgtttgcaccgcaagaagctggcttcccatggcttcccatg 1680

Qy      1681 cgccaccggcttgccgccgcgatgagctgcgtggttcgcggcaggaccacattaaggaatcc 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1681 cgccaccggcttgccgccgcgatgagctgcgtggttcgcggcaggaccacattaaggaatcc 1740

Qy      1741 ggtactgaagatatctcaacaagcaggctgctggatatgctgaacgagcaccgcgcgat 1800
```

```
Db   1741 ggtactgaagatatcttcaacaagcaggcctgctggactgtccatttatggtgtggcacggt 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1801 ggcgtcagatcattcccgtcgactggccgttctgtcatttatggtgtggcacggc 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 ggcgtcagatcattcccgtcgactggccgttctgtcatttatggtgtggcacggc 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1861 atttttgtggaaaaccgcattgatccacagattgaggaccgctcctacccggtcgagctt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 atttttgtggaaaaccgcattgatccacagattgaggaccgctcctacccggtcgagctt 1920
```

RESULT   2
AAA10293
ID   AAA10293 standard; DNA; 3825 BP.
XX
AC   AAA10293;
XX
DT   03-JUL-2000 (first entry)
XX
DE   DNA encoding Corynebacterium glutamicum lysozyme insensitivity protein.
XX
KW   Lysozyme insensitivity protein; bacterium; recombinant expression;
KW   amino acid production; ds.
XX
OS   Corynebacterium glutamicum.
XX
FH   Key   Location/Qualifiers
FT   CDS   815..2737
FT   /tag= a
FT   /product= "Corynebacterium glutamicum lysozyme
FT   insensitivity protein"
XX
PN   WO200014241-A1.
XX
PD   16-MAR-2000.
XX
PF   04-SEP-1998;   98WO-JP03981.
XX
PR   04-SEP-1998;   98WO-JP03981.
XX
PA   (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI   Nagai K, Wati M;
XX
PR   WPI; 2000-256989/22.
     P-PSDB: AAY87459.
XX
PT   New DNA molecule useful for production of amino acids encodes protein
PT   imparting lysozyme resistance to Corynebacterium glutamicum strain -
XX
PS   Example 2; Page 40-44; 50pp; Japanese.
XX
CC   This sequence represents DNA encoding a Corynebacterium glutamicum
CC   lysozyme insensitivity protein. This protein renders lysozyme-sensitive
CC   strains of Corynebacterium glutamicum insensitive to lysozyme. DNA
CC   sequences encoding the lysozyme insensitivity protein can be used to
CC   construct a recombinant vector for the expression of this protein in a
CC   host cell. The DNA encoding the lysozyme insensitivity protein can also
CC   be mutated, and used to generate strains of Corynebacterium glutamicum in
CC   which this protein is inactivated and which are therefore lysozyme-
CC   sensitive. Corynebacterium glutamicum is used for the production of amino
CC   acids, in particular glutamic acid and glutamine.
XX
SQ   Sequence 3825 BP; 881 A; 1031 C; 993 G; 920 T; 0 other;
```

Query Match     100.0%;  Score 1920;  DB 21;   Length 3825;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 1920;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Db   Sequence 3825 BP; 881 A; 1031 C; 993 G; 920 T; 0 other;
```
Qy   1 atgtgcggcctccttgcgcatattgactgactgcaaatgggaacgctgaagcattcgttcctgca 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    815 atgtgcggcctccttgcgcatattgactgactggccaccgtgtcctgtgtggaacgctgaagcattcgttcctgca 874
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 ctcgagcgggccttgccatgccatgcgtcgactggccaccgtgtcctgacgatggccgcacttgcat 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    875 ctcgagcgggccttgccatgcgatgcgtcgactggccaccgtggtcctgacgatgccgcacttgcat 934
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    121 gacgccgatgcagcgtttggattcaacgcgccctccatcattgatattgcacactccac 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    935 gacgccgatgcagcgtttggattcaacgcgccctccatcattgatattgcacactcccac 994
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    181 caaccactgcgtcgggacctgcggatgaaccgaccgcaccgcaccgcacgctcaacgt 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    995 caaccactgcgtcgggaccctgcggatgaaccgaccgcaccgcaccgcacgctcaacgt 1054
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    241 gagatcacactacgtgactcgtgctaaaggctctcgatttggatataccctttaat 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1055 gagatcacactacgtgagctcgtgctaaaggctctcgatttgggatataccctttaat 1114
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    301 acttctggcgatggcgagcaattgtcgttccaccactgggcgagtccggtc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1115 acttctggcgatggcgagcaattgtcgttcgttccaccactgggcgagtccggtc 1174
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    361 gagcatctcccggaatgttcggcattgcattgggatacaaagaaagtcgctttc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1175 gagcatctcccggaatgttcggcattgcattgggatacaaagaaagtcgctttc 1234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    421 cttgcgtgatcagttcggcatcaagcactgtctacgcaccaccgagcatggcacc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1235 cttgcgtgatcagttcggcatcaagccactgtctacgcaccaccgagcatggcacc 1294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    481 gtgttcctccagagagaagaccatctggaagtgccgagagtgaatctagatctg 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1295 gtgttcctccagagagaagaagaccatctggaagtgccgagagtgaatctagatctg 1354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    541 ggccttgataagcgcaccattgagcactgagctgacctgcgtcagtgctgccagat 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1355 ggccttgataagcgcaccattgagcactgagctgacctgcgtcagtgctgccagat 1414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    601 acccttcacgcgcagattcccgctggagtcaggctcgtcagcagttcgtccgggc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1415 acccttcacgcgcagattcccgctggagtcaggctcgtcagcagttcgtccgggc 1474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    661 ggcaagctggaacagaagcgttacttcaagcctcagttcccagtacagaaggtcgtaaag 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1475 ggcaagctggaacagaagcgttacttcaagcctcagttcccagtacagaaggtcgtaaag 1534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    721 ggtaaggagcaggacgacctcttcgatcgcattcgcatcgcactgccccaggtgttggagagcgggaaag 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1535 ggtaaggagcaggacgacctcttcgatcgcattcgcatcgcactgccccaggtgttggagagcgggaaag 1594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    781 catatcgtgccgacgtgacgtagctcgtccttccgggcgcattgactcaaccgca 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1595 catatcgtgccgacgtgacgtaggctcgtccttccgggcgcattgactcaaccgca 1654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    841 attgcgcccgcttgcaaagcgccacaaccctgacctgctcacctcaccccacccggttcgag 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1655 attgcgccgcttgcaaagcgccacaaccctgacctgctcacctcacctcaccccacccggttcgag 1714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    901 cgtgaaggctactcggaggtcgatggctcggagtccgcgctcgcgattggcgctgag 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1715 cgtgaaggctactcggaggtcgatggctcggagtccgcgctcgcgattggcgctgag 1774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    961 cacatcgtgaagattgctcgcctgaggaatacgcaacgcgattcctaagatcatgtgg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1775 cacatcgtgaagattgctcgcctgaggaatacgcaacgcgattcctaagatcatgtgg 1834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1021 tacttggatgatctgtagctgaccacatcattgtccgctgcacttcgtggcagcggaa 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1835 tacttggatgatctgtagctgaccacatcattggtccgctgtactcgtggcagcggaa 1894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1081 gcacgtaagcacgtcaagttgtgtctgcgcagggcgcagagtagctgttcggtgga 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1895 gcacgtaagcacgtcaagttgtgtctgtctgcgcagggcgcagagaagctcggtgga 1954
```

```
PF    18-DEC-2000; 2000EP-0127688.
XX
PR    16-DEC-1999;  99JP-0377484.
PR    07-APR-2000; 2000JP-0159162.
PR    03-AUG-2000; 2000JP-0280988.
XX
PA    (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI    Nakagawa S,  Mizoguchi H, Ando S,  Hayashi M,  Ochiai K,  Yokoi H;
PI    Tateishi N,  Senoh A,  Ikeda M,  Ozaki A;
XX
DR    WPI; 2001-376931/40.
DR    P-PSDB; AAG92152.
XX
PT    Novel polynucleotides derived from Coryneform bacteria, for identifying
PT    mutation point of a gene, measuring expression of a gene, analysing
PT    expression profile or pattern of a gene and identifying homologous gene
PT    -
XX
PS    Claim 8; SEQ ID NO: 2406; 246pp + Sequence Listing; English.
XX
CC    The present invention provides a number of nucleotide and protein
CC    sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC    are useful for identifying the mutation point of a gene derived from a
CC    mutant of coryneform bacterium, measuring expression amount and
CC    analysing the expression profile or expression pattern of a gene derived
CC    from Coryneform bacterium, and identifying a homologue of a gene derived
CC    from coryneform bacterium. Coryneform bacteria are useful for producing
CC    amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC    particularly L-lysine. The present sequence is a nucleic acid described
CC    in the exemplification of the invention.
CC    Note:The sequence data for this patent did not form part of the printed
CC    specification, but was obtained in electronic format directly from the
CC    European Patent Office.
XX
SQ    Sequence 1920 BP; 422 A; 534 C; 528 G; 436 T; 0 other;

Query Match             99.2%;  Score 1904;  DB 22;  Length 1920;
Best Local Similarity   99.5%;  Pred. No. 0;
Matches 1910;  Conservative   0;  Mismatches   10;  Indels   0;  Gaps   0;

Qy      1 atgtgcggcctcttgcatatgactgcaaatggcaacgctgaagcattcgttcctgca 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 atgtgcggcctcttggcatatgactgcaaatggcaacgctgaagcattcgttcctgca 60

Qy     61 ctcgagcgggccttgcccatgcatgcgcaccgtggtcctgacgatgcgcacttggcat 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ctcgagcgggccttgcccatgcatgcgcaccgtggtcctgacgatgcgcacttggcat 120

Qy    121 gacgccgatgcagcgtttgggacctcggatgaaccgacccgctacgcacggtttcaacggt 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 gacgccgatgcagcgtttgggatccaaccgctcgatattgcacactcccac 180

Qy    181 caaccactgcgttggggacctcggatgaaccgacccgctacgcaatgactttcaacggt 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 caaccactgcgttggggacctcggatgaaccgacccgctacgcaatgactttcaacggt 240

Qy    241 gggatctcaactacgttgagctgcgtaaagagctctcggatttggattataccttaat 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 gagatctcaactacgttgagctgcgtaaagagctctcggatttggatatgccttaat 300

Qy    301 acttctgcgatgcggagccaatgttgttgcggttccaccactgttgcggcactggcat 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 acttctgcgatgcggagccaatgttgtccggttccaccactggggagtccgtggtc 360

Qy    361 gagcatcccgacgaatgttcggcattgccattggatacaaaggaaaagtcgctttcc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 gagcatcccgacgaatgttcggcattggattggatacaaaggaaaagtcgctttcc 420

Qy    421 cttgcgcgtgatcagttcggcatcaagcacgttctacgcaaccaccagcatggcacc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 cttgcgcgtgatcagttcggcattaagcacgttctacgcaaccaccagcatggcacc 480
```

```
Qy   1141 tacaccattacaaagagcgctatcgcttgctccatttgagagatccctcccacta 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1955 tacaccattacaaagagcgctatcgcttgctccatttgagagatccctcccacta 2014

Qy   1201 cgtaaaggcctggaaagctcagcaaggttctgccagacggcatgaagggcaagtccctt 1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2015 cgtaaaggcctggaaagctcagcaaggttctgccagacggcatgaagggcaagtccctt 2074

Qy   1261 cttgagcgtgtggctccatgacgatgaagagcgctactacggcaacgctcgctccttcaat 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2075 cttgagcgtggctccatgacgatgaagagcgctactacggcaacgctcgctccttcaat 2134

Qy   1321 ttcgagcagatgcaacgcgttattccatggccaaagcgcgatggaccaccgcgaagtc 1380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2135 ttcgagcagatgcaacgcgttattccatggccaaagcgcgatggaccaccgcgaagtc 2194

Qy   1381 actgcaccgatctacgcacacaatcccgcaactttgatccagtagcccgcatgcaacacctg 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2195 actgcaccgatctacgcacacaatcccgcaacttttgatccagtagcccgcatgcaacacctg 2254

Qy   1441 gatctgttcacctggatgcgcgggcgacatcctggtcaaggctgacaacatggcg 1500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2255 gatctgttcacctggatgcgcgggcgacatcccggtcaaggctgacaagatcaacatggcg 2314

Qy   1501 aactccctggagctgcagtccagtcgagttccatctctgataaggaagtttcaaggttgcagagacc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2315 aactccctggagctgcagtccagtcgagttccatctctgataaggaagtttcaaggttgcagagacc 2374

Qy   1561 attccttacgatctgaagattgccaacgtaccaccagtacggctgcgcgcaggggcactc 1620
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2375 attccttacgatctgaagattgccaacgtaccaccaagtacggctgcgcgcaggggcactc 2434

Qy   1621 gagcagattgttccggctcacgtttgcacgcaagaagctgggcttccctgttcccatg 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2435 gagcagattgttccggctcacgtttgcacgcaagaagctggcttccctgttcccatg 2494

Qy   1681 cgccactgcgcttgccggcgcgatgagagctgttcggttggggcgcaggacaccattaaggaatcc 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2495 cgccactggctgccggcgcgatgagagctgttcggttgggcgcaggacaccattaaggaatcc 2554

Qy   1741 ggtactgaagatatcttcaacagcaggctgctgctgatatctgaacgagcacccgat 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2555 ggtactgaagatatcttcaacagcaggctgctgctgatatctgaacgagcacccgat 2614

Qy   1801 ggcgtgtcagatcattcccgtcgactgtgactgttctgtcatttatggtgtggcacggc 1860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2615 ggcgtgtcagatcattcccgtcgactgtgactgttctgtcatttatggtgtggcacggc 2674

Qy   1861 atttttgtggaaaaccgcattgatccacagattgaggacccgctccaccggtcgagctt 1920
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2675 atttttgtggaaaaccgcattgatccacagattgaggacccgctccaccggtcgagctt 2734
```

```
RESULT  3
AAH67371
ID     AAH67371 standard; DNA; 1920 BP.
XX
AC     AAH67371;
XX
DT     26-SEP-2001  (first entry)
XX
DE     C glutamicum coding sequence fragment SEQ ID NO: 2406.
XX
KW     Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW     organic acid synthesis; ds.
XX
OS     Corynebacterium glutamicum.
XX
PN     EP1108790-A2.
XX
PD     20-JUN-2001.
XX
```

Qy 1561 attccttacgatctgaagattgccaaccggtaccaccagtacgcgctgcgcaggcactc 1620
         ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1561 attccttacgacctgaagattgccaacggtaccaccagtacgcgctgcgcaggcactc 1620

Qy 1621 gagcagattgttccgctcacgtttgcaccgcaagaagctgggcttcccgttcccatg 1680
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1621 gagcagattgttccgctcacgtttgcaccgcaagaagctgggcttcccgttcccatg 1680

Qy 1681 cgccactggcttgccggcgatgagctgtcggtcggttgggcgcaggacaccattaaggaatcc 1740
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1681 cgccactggcttgccggcgatgagctgtcggtcggttgggcgcaggacaccatcaaggaatcc 1740

Qy 1741 ggtactgaagatatctcaacaagcagcgctgtgctggatatgtgaacgagcaccgcgat 1800
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1741 ggtactgaagatatctcaacaagcagcgctgtgctggatatgtgaacgagcaccgcgcat 1800

Qy 1801 ggcgtgtcagatcattcccgtcgactgtggactgttctgtcattatggtgtggcacggc 1860
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1801 ggcgtgtcagatcattcccgtcgactgtggactgttctgtcattatggtgtggcacggc 1860

Qy 1861 atttttgtggaaaaccgcattgatccacagattgaggaccgctcccaccagtcgagctt 1920
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1861 atttttgtggaaaaccgcattgatccacagattgaggaccgctcccaccagtcgagctt 1920

RESULT   4
AAH68531

ID   AAH68531 standard; DNA; 349980 BP.
XX
AC   AAH68531;
XX
DT   26-SEP-2001 (first entry)
XX
DE   C glutamicum coding sequence fragment SEQ ID NO: 7066.
XX
KW   Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW   organic acid synthesis; ds.
XX
OS   Corynebacterium glutamicum.
XX
PN   EP1108790-A2.
XX
PD   20-JUN-2001.
XX
PF   18-DEC-2000; 2000EP-0127688.
XX
PR   16-DEC-1999; 99JP-0377484.
PR   07-APR-2000; 2000JP-0159162.
PR   03-AUG-2000; 2000JP-0280988.
XX
PA   (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI   Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI   Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR   WPI; 2001-376931/40.
XX
PT   Novel polynucleotides derived from Coryneform bacteria, for identifying
PT   mutation point of a gene, measuring expression of a gene, analysing
PT   expression profile or pattern of a gene and identifying homologous gene
PT   -
XX
PS   Disclosure; SEQ ID NO: 7066; 246pp + Sequence Listing; English.
XX
CC   The present invention provides a number of nucleotide and protein
CC   sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC   are useful for identifying the mutation point of a gene derived from a
CC   mutant of coryneform bacterium, measuring expression amount and
CC   analysing the expression profile or expression pattern of a gene derived
CC   from Coryneform bacterium, and identifying a homologue of a gene derived
CC   from coryneform bacterium. Coryneform bacteria are useful for producing
CC   amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC   particularly L-lysine. The present sequence is a nucleic acid described

Qy 481 gtgttcctctcagagaagaagaccatcttggagatggccgagagagtgaatctagatctg 540
         ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db 481 gtgttcctctcagagaagaagaccatcttggagatggccgagagagtgaatctagatctg 540

Qy 541 ggccttgataagcgcaccattgagcactgagcactgcagtacgcagtacgtgcccgagccagat 600
         ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db 541 ggccttgataagcgcaccattgagcactgagcactgcagtacgtgcagtacgtgcccgagccagat 600

Qy 601 acccttcacgcgccagatttcccgcttggagtcaggctgacccgcaacagttcgtccgggc 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601 acccttcacgcgccagatttcccgcttggagtcaggctgacccgcaacagttcgtccgggc 660

Qy 661 ggcaagctggaacagaaggcgttactcaagcctcagttcccagtacagaaggtcgtaaag 720
         |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db 661 ggcaagctggaacagaaggcgttactcaagcctcagttcccagtacagaaggtcgtaaag 720

Qy 721 ggtaaggagcaggacccctcttcgattgccaggtgttggaggatagcgtcgaaaag 780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 721 ggtaaggagcaggacccctcttcgattgccaggtgttggaggatagcgtcgaaaag 780

Qy 781 catatgcgtgccgacgtgaccgtaggctcgttccttccggcgcattgactcaaccgca 840
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 781 catatgcgtgccgacgtgaccgtaggctcgttccttccggcgcggcattgactcaaccgca 840

Qy 841 atggcgccgcttgcaaagcgccacaacccctgacctgctcaccttcaccacggtttcgag 900
         ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db 841 atggcggccgcttgcaaagcgccacaacccctgacctgctcacccttcaccacccggtttcgag 900

Qy 901 cgtgaaggctactcggaggtcgatgtgtggcgagtccgcgcgattggcgctgag 960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 901 cgtgaaggctactcggaggtcgatgtgtggcgagtccgcgcgattggcgctgag 960

Qy 961 cacactgtgaagattgtctccgcctgaggaatacgccaacgcgattcctaagatcatgtgg 1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 961 cacactgtgaagattgtctccgcctgaggaatacgccaacgcgattcctaagatcatgtgg 1020

Qy 1021 tacttggatgatctgtagctgaccatcattgtccgctacttcgtggcagcggaa 1080
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 tacttggatgatctgtagctgaccatcattgtccgctacttcgtggcagcggaa 1080

Qy 1081 gcacgtaagcacgtcaaggttgtctgctggcgaggcgcagatgagctgttcggtgga 1140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 gcacgtaagcacgtcaaggttgtctgctggcgaggcgcagatgagctgttcggtgga 1140

Qy 1141 tacaccattacaaagagccgctatcgctctgctcctccccacta 1200
         ||||||||||||||||||||||||||||||||||||||||||||
Db 1141 tacaccattacaaagagccgctatcgctctgctcctccccacta 1200

Qy 1201 cgtaaaggctgggaaagctcagcaaggtctgccagacggcatgaagggcaagtcccctt 1260
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201 cgtaaaggctgggaaagctcagcaaggtctgccagacggcatgaaggggcaagtcccctt 1260

Qy 1261 cttgagcgtggcctccatgacaccatggaagagcgctactacggcaacgctcttcttcaat 1320
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1261 cttgagcgtggcctccatgaccaccatggaagagcgctactacggcaacgctcgctcctccttcaat 1320

Qy 1321 ttcgagcagatgcaacgcgttattccactggcaaagcgcgaagtggacaccgcgaagtc 1380
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1321 ttcgagcagatgcaacgcgttattccactggcaaagcgcgaatgggaccaccgcgaagtc 1380

Qy 1381 actgcaccgatctacgcacaacaatccgcaactgatccagtacgccgatgcaacacctg 1440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1381 actgcaccgatctacgcacaacaatccgcaactgatccagtacgccgatgcaacacctg 1440

Qy 1441 gatctgttcacctgatgcgcggcgacatcaaagctgacaagatcaacatgcg 1500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1441 gatctgttcacctgatgcgcggcgacatcaaagctgacaagatcaacatggcg 1500

Qy 1501 aactccctgagctcgcgagttccattcttggataaggaagtttcaaggttgcagagacc 1560
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1501 aactccctgagctcgcgagttccattcttggataaggaagtttcaaggttgcagagacc 1560

```
CC    in the exemplification of the invention.
CC    Note: The sequence data for this patent did not form part of the printed
CC    specification, but was obtained in electronic format directly from the
CC    European Patent Office.
XX
SQ    Sequence 349980 BP; 80724 A; 98367 C; 90490 G; 80399 T; 0 other;

Query Match    99.2%;  Score 1904;  DB 22;  Length 349980;
Best Local Similarity 99.5%;  Pred. No. 0;
Matches 1910;  Conservative 0;  Mismatches 0;  Indels 10;  Gaps 0;

Qy      1 atgtgcggccttcttggcatattgactgcatatgaaatggaaccgctgaagcattcgttcctgca 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228516 atgtgcgggccttcttggcatattgactgcatatgaaatggaaccgctgaagcattcgttcctgca 228575

Qy     61 ctcgagcgggccttgccatgcatgcgatgccggcactggcat 120
          |||||||||||||||||||||||||||||||||||||||
Db 228576 ctcgagcgggccttgccatgccatgcgatgccggcactggcat 228635

Qy    121 gacgccgatgcgagcgtttggattcaaccgcctctccatcattgatattgcacactcccac 180
          |||||||||||||||||||||||||||||||||||||||||||||
Db 228636 gacgccgatgcgagcgtttggattcaaccgcctctccatcattgatattgcacactcccac 228695

Qy    181 caaccactggttggggacctgcggatgaacccgaccgcctacggcaatgactttcaacggt 240
          |||||||||||||||||||||||||||||||||||||||||||||
Db 228696 caaccactgcggtggggacctgcggatgaaccgcgaccgcctacggcaatgactttcaacggt 228755

Qy    241 gagatctacaactacgttgagctgactggctaaagagctctcggattgggatatacctttaat 300
          ||||||||||||||||||||||||||||||||||||||||||||
Db 228756 gagatctacaactacgttgagctgactgggctaaagagctctcggattgggatatgcctttaat 228815

Qy    301 actttctggcgatggcgagccaattgttgtcggttccacctgggcgatcggtggtc 360
          |||||||||||||||||||||||||||||||||||||||||||
Db 228816 actttctggcgatggcgagccaattgttgtcggttccacctgggcgagtccggtggtc 228875

Qy    361 gagcatctccgcggaatgttcggcattgccattgggatacaaagggaaaagtcgctttttc 420
          |||||||||||||||||||||||||||||||||||||||||||||
Db 228876 gagcatctccgcggaatgttcggcattgccattgggatacaaaggaaaaagtcgctttttc 228935

Qy    421 cttgcgcgtgatcagttcggcatcaagccactgttctacgcaccaccgagcatggcacc 480
          |||||||||||||||||||||||||||||||||||||||||||
Db 228936 cttgcgcgtgatcagttcggcatcaagccactgttctacgcaccaccgagcatggcacc 228995

Qy    481 gtgttccctcagaagaagaaccatctggagatgcgcgaggagatggaattctagatctg 540
          |||||||||||||||||||||||||||||||||||||||||||
Db 228996 gtgttccctcagaagaagaaccatctggagatgcgcgaggagatggaatctagatctg 229055

Qy    541 ggccttgataagcgcaccattgagcactacggtgacctgctcagtgccgagccagat 600
          |||||||||||||||||||||||||||||||||||||||||||
b 229056 ggccttgataagcgcaccattgagcactacggtgacctgctcagtgccgagccagat 229115

Qy    601 acccttcacgcgcagaattcccgcttggagtcaggctgcgctcagagtcgtcggc 660
          ||||||||||||||||||||||||||||||||||||||||||
Db 229116 acccttcacgcgcagaattcccgcttggagtcaggctgcgctcagagtcgtcgggc 229175

Qy    661 ggcaagctggaacagtggaacgttacttcaagctgcagagtacagaaggtcgtaaag 720
          |||||||||||||||||||||||||||||||||||||||||||
Db 229176 ggcaagctggaactgaacgttacttcaagctgcagagtacagaaggtcgtaaag 229235

Qy    721 ggtaaggaggaggaccctgccaccgattgtcggagatagcgtcgaaaag 780
          ||||||||||||||||||||||||||||||||||||||||
Db 229236 ggtaaggaggaggaccctgccaccgattgtcggagatagcgtcgaaaag 229295

Qy    781 catatgtgccgccgacctgaccgtagcgtaggctgttccttccggcgcattgactcaaccgca 840
          |||||||||||||||||||||||||||||||||||||||||||||
Db 229296 catatgtgccgccgacctgaccgtagcgtaggctgttccttccggcgcattgactcaaccgca 229355

Qy    841 attgcgccgcttgcaaagcgccacaaccctgacctgctcacctcaccggtttcgag 900
          ||||||||||||||||||||||||||||||||||||||||||
Db 229356 attgcgccgcttgcaaagcgccacaaccctgacctgctcacctcaccggtttcgag 229415

Qy    901 cgtgaaggctactcggaggtcgatgtggctgcggaccgcgctgcgattggcgctgag 960


Db 229416 cgtgaaggctactcggaggtcgatgtggctgcgcgactgcgccgctgccgattggcgctgag 229475
          |||||||||||||||||||||||||||||||||||||||||||||
Qy    961 cacatcgtgaaagattgctctcgccgcgaggaataccgccaaccggatcctaagatcatgtgg 1020
          ||||||||||||||||||||||||||||||||||||||||||||
Db 229476 cacatcgtgaagattgctctcgccgcgaggaataccgccaaccgatcctaagatcatgtgg 229535

Qy   1021 tacttggatgatctgtagctgaccaccatcggtcccgctgtactcgtcggcgggaa 1080
          |||||||||||||||||||||||||||||||||||||||||||
Db 229536 tacttggatgatctgtagctgaccaccatcggtcccgctgtactcgtcggcggaa 229595

Qy   1081 gcacgtaagcacgtcaagttgtgtctgtctcggcgagggcgcagatgagcgttcggtgga 1140
          |||||||||||||||||||||||||||||||||||||||||||
Db 229596 gcacgtaagcacgtcaagttgtgtctgtctcggcgaggggcgcagatgagcgttcggtgga 229655

Qy   1141 tacaccattacaaagacgcgtatccgtctgtccattgaaagatccctcccacta 1200
          ||||||||||||||||||||||||||||||||||||||||||
Db 229656 tacaccattacaaagacgcgtatccgtctgtccattgagaagatccctcccacta 229715

Qy   1201 cgtaaagcctgggaaagctcagcaaggttctgccagacgcatgaaggcaagtcccct 1260
          |||||||||||||||||||||||||||||||||||||||||||
Db 229716 cgtaaagcctgggaaagctcagcaaggttctgccagacggcatgaaggcaagtcccct 229775

Qy   1261 cttgagcgtggctccatgaccatggaagagcgctactacggcaacgctcgctccttcaat 1320
          ||||||||||||||||||||||||||||||||||||||||||||
Db 229776 cttgagcgtggctccatgaccatggaagagcgctactacggcaacgctcgctccttcaat 229835

Qy   1321 tcgagcagatcaacgcgttattccatgggcaaagccgaatggcaagcgcgaagtc 1380
          |||||||||||||||||||||||||||||||||||||||||
Db 229836 ttcgagcagatcaacgcgttattccatgggcaaagccgaatggcaagcgcgaagtc 229895

Qy   1381 actgcaccgatctacgcacaatcccgcaacttgatccagtagccgcatgcaacacctg 1440
          ||||||||||||||||||||||||||||||||||||||||||
Db 229896 actgcgccgatctacgcacaatcccgcaacttgatccagtagcccgcatgcaacacctg 229955

Qy   1441 gatctgttcaccggatgcgcggcgacatcctgtcaagctgacaagatcaacatgtgcg 1500
          |||||||||||||||||||||||||||||||||||||||||||
Db 229956 gatctgttcaccggatgcgcgcgcacatcctgtcaagctgacaagatcaacatggcg 230015

Qy   1501 aactccttgactgcgagttccattcttggataagaagtttcaaggttgcagagacc 1560
          ||||||||||||||||||||||||||||||||||||||||||
Db 230016 aacccctgactgcgagttccattcttggataagaagtttcaaggttgcagagacc 230075

Qy   1561 attcctacgatctgaagatgccaccagagtaccagcgtctgcgcaggggcactc 1620
          |||||||||||||||||||||||||||||||||||||||||
Db 230076 attcctacgatctgaagatgccaccagagtaccagcgtctgcgcaggggcactc 230135

Qy   1621 gagcagattgttccgcctcacgtttgcaccgcaagagctggcttcctgttcccatg 1680
          |||||||||||||||||||||||||||||||||||||||||||
Db 230136 gagcagattgttccgcctcacgtttgcaccgcaagagctggcttcctgttcccatg 230195

Qy   1681 cgccactggctgcgcgatagcgtcggttggcgcaggacaccattaaggaatcc 1740
          ||||||||||||||||||||||||||||||||||||||||||
Db 230196 cgccactggctgcgcgatagcgtcggttggcgcaggacaccatcaaggaatcc 230255

Qy   1741 ggtactgaagatatctcaacagcaggctgctgctgatatgcgaacgagccgcgat 1800
          |||||||||||||||||||||||||||||||||||||||||||
Db 230256 ggtactgaagatatctcaacagcaggctgctgctgatatgcgaacgagccgcgat 230315

Qy   1801 ggcgtgtcagatcattccgctcgactgttctgtcatttatggtggcacggc 1860
          |||||||||||||||||||||||||||||||||||||||||
Db 230316 ggcgtgtcagatcattccgctcgactgttctgtcatttatggtggcacggc 230375

Qy   1861 attttttggaaaaccgccattgatccacagattgagcacccgtcgagctt 1920
          ||||||||||||||||||||||||||||||||||||||||||
Db 230376 attttttggaaaaccgccattgatccacagattgagcacccgtcgagctt 230435

RESULT 5
AAF71807
ID   AAF71807 standard; DNA; 1962 BP.
XX
AC   AAF71807;
```

```
XX
DT   30-APR-2001  (first entry)
XX
DE   Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:109.
XX
KW   Corynebacterium glutamicum; metabolic pathway protein; MP protein;
KW   fine chemical production; microorganism; organic acid; nucleoside;
KW   nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
KW   lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
KW   carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.
XX
OS   Corynebacterium glutamicum.
XX
PN   WO200100843-A2.
XX
PD   04-JAN-2001.
XX
PF   23-JUN-2000; 2000WO-IB00923.
XX
PR   25-JUN-1999; 99US-0141031.
PR   01-JUL-1999; 99DE-1030476.
PR   02-JUL-1999; 99US-0142101.
PR   08-JUL-1999; 99DE-1031415.
PR   08-JUL-1999; 99DE-1031418.
PR   08-JUL-1999; 99DE-1031419.
PR   08-JUL-1999; 99DE-1031420.
PR   08-JUL-1999; 99DE-1031424.
PR   08-JUL-1999; 99DE-1031428.
PR   08-JUL-1999; 99DE-1031434.
PR   08-JUL-1999; 99DE-1031435.
PR   08-JUL-1999; 99DE-1031443.
PR   08-JUL-1999; 99DE-1031453.
PR   08-JUL-1999; 99DE-1031457.
PR   08-JUL-1999; 99DE-1031465.
PR   08-JUL-1999; 99DE-1031478.
PR   08-JUL-1999; 99DE-1031510.
PR   08-JUL-1999; 99DE-1031541.
PR   08-JUL-1999; 99DE-1031573.
PR   08-JUL-1999; 99DE-1031592.
PR   08-JUL-1999; 99DE-1031632.
PR   08-JUL-1999; 99DE-1031634.
PR   08-JUL-1999; 99DE-1031636.
PR   09-JUL-1999; 99DE-1032125.
PR   09-JUL-1999; 99DE-1032126.
PR   09-JUL-1999; 99DE-1032130.
PR   09-JUL-1999; 99DE-1032186.
PR   09-JUL-1999; 99DE-1032206.
PR   09-JUL-1999; 99DE-1032227.
PR   09-JUL-1999; 99DE-1032228.
PR   09-JUL-1999; 99DE-1032229.
PR   09-JUL-1999; 99DE-1032230.
PR   14-JUL-1999; 99DE-1032922.
PR   14-JUL-1999; 99DE-1032926.
PR   14-JUL-1999; 99DE-1032928.
PR   14-JUL-1999; 99DE-1033004.
PR   14-JUL-1999; 99DE-1033005.
PR   09-JUL-1999; 99DE-1033006.
PR   12-AUG-1999; 99US-0148613.
PR   27-AUG-1999; 99DE-1040764.
PR   27-AUG-1999; 99DE-1040765.
PR   27-AUG-1999; 99DE-1040766.
PR   27-AUG-1999; 99DE-1040832.
PR   31-AUG-1999; 99DE-1041378.
PR   31-AUG-1999; 99DE-1041379.
PR   31-AUG-1999; 99DE-1041380.
PR   31-AUG-1999; 99DE-1041394.
PR   31-AUG-1999; 99DE-1041396.
PR   03-SEP-1999; 99DE-1042076.
PR   03-SEP-1999; 99DE-1042077.
PR   03-SEP-1999; 99DE-1042079.
PR   03-SEP-1999; 99DE-1042086.
PR   03-SEP-1999; 99DE-1042087.
PR   03-SEP-1999; 99DE-1042088.
PR   03-SEP-1999; 99DE-1042095.
PR   03-SEP-1999; 99DE-1042124.
PR   03-SEP-1999; 99DE-1042129.
PR   09-MAR-2000; 2000US-0187970.
XX
PA   (BADI ) BASF AG.
XX
PI   Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
DR   WPI; 2001-137957/14.
DR   P-PSDB; AAB79688.
XX
PT   Nucleic acids from Corynebacterium glutamicum encoding metabolic
PT   pathway proteins, useful for producing fine chemicals in
PT   microorganisms, including organic acids, nonproteinogenic amino acids,
PT   and purine and pyrimidine bases -
XX
PS   Claim 3; Page 324-327; 1737pp; English.
XX
CC   AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
CC   pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
CC   MP nucleic acids are useful for the production of fine chemicals
CC   in microorganisms, including organic acids, nonproteinogenic amino
CC   acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
CC   saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
CC   compounds, vitamins, cofactors, polyketides and enzymes.
XX
SQ   Sequence 1962 BP; 435 A; 544 C; 534 G; 449 T; 0 other;

Query Match             99.1%;  Score 1902.4;  DB 22;  Length 1962;
Best Local Similarity   99.4%;
Matches 1909;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          1 atgtgcggcctcttggcatattgactgcaatggaacgctgaacattcgttcctgca 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         20 atgtgcggcctcttggcatattgactgcaatgggaacgctgaacattcgttcctgca 79

Qy         61 ctcgagcgggccttgccatgcatgcgccaccgtgtcctgacgatgcgccacttggcat 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         80 ctcgagcgggccttgccatgcatgcgccaccgtgtcctgacgatgcgccacttggcat 139

Qy        121 gacgccgatgcagcgtttgattcaaccgcctccccatcattgatattgcacactcccac 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        140 gacgccgatgcagcgtttgattcaaccgcctcctccatcattgatattgcacactcccac 199

Qy        181 caaccactgcgtggggacctgcggatgaaccgcggcaatgctacgcaatgactttcaacggt 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        200 caaccactgcgtggggacctgcggatgaaccgcggcaatgctacgcaatgactttcaacggt 259

Qy        241 gagatctacaactacgttgagctgcgtaaagagctctcggatttggatataccctttaat 300
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        260 gagatctacaactacgttgagctgcgtaaagagctctcggatttgggatatgcctttaat 319

Qy        301 acttctggcgatggcggagccagtgttcgtcggttccaccactcgggcgagtccgtggtc 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        320 acttctggcgatggcggagccagtgttcgtcggttccaccactcgggcgagtccgtggtc 379

Qy        361 gagcatccgcggaatgttcggcattgccatttgggatacaaaggaaaagtcgcttttc 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 gagcatccgcggaatgttcggcattgccatttgggatacaaaggaaaagtcgcttttc 439

Qy        421 cttgcgcgtgatcagttcggcatcaagcccgagcaaccacccgagcatggcacc 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 cttgcgcgtgatcagttcggcattaagcccgagcaaccacccgagcatggcacc 499

Qy        481 gtgttcctcagagagaagaccatcttggagatgtgccgagagatgaatctagatctg 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 gtgttcctcagagagaagagaccatcttggagatgtgccgagagatgaatctagatctg 559

Qy        541 ggcctgataagcgcaccattgagcactacgtgacctgcagtacgtgccgagccagat 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 ggcctgataagcgcaccattgagcactacgtgacctgcagtacgtgccgagccagat 619
```

```
Qy   601 accctcacgcgcagattcccgctggagtcaggctgcgaccgcaacagttcgtccggc 660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   620 accctcacgcgcagattcccgctggagtcaggctgcgaccgcaacagttcgtccggc 679

Qy   661 ggcaagctggaacagagcgttacttcaagctcagttccagtacagagaggtcgtaaag 720
         |||||||||||||||||||||||||||||||||||||||||||||
Db   680 ggcaagctggaacagagcgttacttcaagctcagttccagtacagagaggtcgtaaag 739

Qy   721 ggtaaggagcaggagacctcttcgatcgcattgggaccagtgttggaggatagcgtcgaaaag 780
         ||||||||||||||||||||||||||||||||||||||
Db   740 ggtaaggagcaggagacctcttcgatcgcattgggaccagtgttggaggatagcgtcgaaaag 799

Qy   781 catatgcgtgccgacgtaggctcgttcctttccgcggccattgactcaaccgca 840
         |||||||||||||||||||||||||||||||||||||||
Db   800 catatgcgtgccgacgtaggctcgttcctttccgcggccattgactcaaccgca 859

Qy   841 atggcgccgctgcaaagcgccacaccctgactcacctcctgctcacctccacggttcgag 900
         ||||||||  ||||||||||||||||||||
Db   860 atggcgggcgcctgcaaagcgccacaccctgactcacctcctgctcacctccacggttcgag 919

Qy   901 cgtgaaggctactcgagggtcgatggtcgtcggagtccgcgcgctcgattggcgctgag 960
         ||||||||||||||||||||||||||||||
Db   920 cgtgaaggcgcctgcaaagcgctcgatgtggctcgtcggagtccgcgcgctcgattggcgctgag 979

Qy   961 cacatcgtgaagatgtctcgctgaggaatacgccaacgcgattcctagatcatgtgg 1020
         |||||||||||||||||||||||||
Db   980 cacatcgtgaagatgtctcgctgaggaatacgccaacgcgattcctagatcatgtgg 1039

Qy  1021 tacttggatgtcctgtagctgagctgtgaccatcattggtccgctgtacttcgtggcacgcgaa 1080
         |||||||||||||||||||||||||
Db  1040 tacttggatgtcctgtagctgagctgtgaccatcattggtccgctgtacttcgtggcacgcgaa 1099

Qy  1081 gcacgtaagcacgtcaaggtcgtcgtgtctggcgaggcgcagatgagctgttcgtgtgga 1140
         |||||||||||||||||||
Db  1100 gcacgtcaagcacgtcaaggtcgtcgtgtctggcgaggcgcagatgagctgttcgtgtgga 1159

Qy  1141 tacaccattacaaagagcgctatcgcttgcttctgctccattgagaagatcccctccccacta 1200
         |||||||||||||||||||||||||||
Db  1160 tacaccattacaaagagcgctatcgcttgcttctgctccattgagaagatcccctccccacta 1219

Qy  1201 cgtaaaggcctgggaaagctcagcaaggttctgcagacgcatgaaggcaagtccctt 1260
         ||||||||||||||||||||||||||
Db  1220 cgtaaaggcctgggaaagctcagcaaggttctgcagacgcatgaaggcaagtccctt 1279

Qy  1261 cttgagcgtggctccatgaccatggaagacgctactacggcaacgctcgctcttcaat 1320
         ||||||||||||||||||||||||||
Db  1280 cttgagcgtggctccatgaccatggaagacgctactacggcaacgctcgctcttcaat 1339

Qy  1321 ttcgagcagatgcaacgcgttattccatggcaaagcgcgaatggaccacgcgaagtc 1380
         ||||||||||||||||||||||||||
Db  1340 ttcgagcagatgcaacgcgttattccatggcaaagcgcgaatggaccacgcgaagtc 1399

Qy  1381 actgcaccgatctacgcacatcccgcaacttgatccagtagcccgcatgcaacacctg 1440
         |||||||||||||||||||||||||||
Db  1400 actgcgccgatctacgcacgtcccgcaacttgatccagtagcccgcatgcaacacctg 1459

Qy  1441 gatctgttcacctgatgcgcggcgacatcctggtcaagctgacaagatcaaacatggcg 1500
         ||||||||||||||||||||||||||
Db  1460 gatctgttcacctggatgcgcggcgacatcctggtcaagctgacaagatcaaacatgggcg 1519

Qy  1501 aactccctggagctgctgcacttctgctccattctggataaggaagtttcaaggttgcagagacc 1560
         ||||||||||||||||||||||||||
Db  1520 aactccctgagctgctgcagttccattctggataaggaagtttcaaggttgcagagacc 1579

Qy  1561 attccttacgatctgaagattgccaacggtaccaccaagtaccgctgcgcaggcactc 1620
         ||||||||||||||||||||||||||
Db  1580 attccttacgacctgaagattgccaacggtaccaccaagtaccgctgcgcaggcactc 1639

Qy  1621 gagcagattgttccgctcacgtttgcaccgcaggtgggcttccctgttcccatg 1680
         ||||||||||||||||||||||
Db  1640 gagcagattgttccgctcacgtttgcaccgcaggtgggctcgggctccgttcccatg 1699
```

```
Qy  1681 cgccactggcttgccggcgatgagcgtgttcggtcggcgcaggagaccaccattaaggaatcc 1740
         |||||||||||||||||||||||||||||||||
Db  1700 cgccactggcttgccggcgatgagcgtgttcggtcggcgcaggacaccatcaaggaatcc 1759

Qy  1741 ggtactgaagatatctcaacaagcaggctgctgctggatatgctgaacgagcaccgcgat 1800
         ||||||||||||||||||||||||||||
Db  1760 ggtactgaagatatctcaacaagcaggctgctgctggatatgctgaacgagcaccgcgat 1819

Qy  1801 ggcgtgtcagatcattccgtcgactgtgaccgttcgtcattatgtgtggcacggc 1860
         |||||||||||||||||||||||||
Db  1820 ggcgtgtcagatcattccgtcgactgtggaccgttcgtcattatgtgtggcacggc 1879

Qy  1861 atttgtggaaaacgcattgatccacagattgagagaccgctcctaccgtcgagctt 1920
         ||||||||||||||||||||||||||
Db  1880 attttgtggaaaacgcattgatccacagattgagagaccgctcctaccgtcgagctt 1939
```

```
Query Match             6.1%;  Score 116.6;  DB 18;  Length 30001;
Best Local Similarity   56.1%;  Pred. No. 2.6e-23;
Matches  243;  Conservative   0;  Mismatches 184;  Indels   6;  Gaps   1;

Qy   81 catgcgcaccgtggtcctgacgatgccgatgacgccggcactggacgttgg 140
        |||   ||| ||  ||   ||| ||  || |  ||   || |||
```

CC The invention relates to combination of two plasmids for cloning the
CC genes encoding the biosynthetic production pathways of chlortetracycline,
CC tetracycline, or their derivatives. One plasmid contains an origin of
CC replication, an actinomycete (such as Streptomyces lividans,
CC S. ariseofuscus, or S. ambofaciens) active antibiotic resistance gene and
CC three or more tandem cohesive end sites. The second plasmid contains an
CC actinomycete active origin of replication and three or more tandem
CC cohesive end sites. The invention can be used to clone a large amount of
CC genetic material for the heterologous production of the antibiotics,
CC chlortetracycline, tetracycline and their analogues. A cluster of genes
CC can be inserted into the plasmid allowing a biosynthetic pathway to be
CC transferred in its entirety to a heterologous host. The invention
CC specifically relates to cloning of the entire tetracycline and
CC chlortetracycline from Streptomyces aureofaciens and its expression in a
CC heterologous host such as S. lividans. The present sequence represents a
CC S. aureofaciens DNA sequence from the cosmid clones designated LP2-127
CC and LP2-128.
XX
SQ Sequence 30001 BP; 4055 A; 10939 C; 11019 G; 3988 T; 0 other;
XX

Query Match   6.1%;   Score 116.6;   DB 20;   Length 30001;
Best Local Similarity 56.1%;   Pred. No. 2.6e-23;
Matches 243;   Conservative   0; Mismatches 184;   Indels   6;   Gaps   1;

Qy   81 catgcgccaccgtgggtcctgacgatgccggcacttggcatgacgccgatgacgccgttgg 140
Db 22834 CATGGCCTGCTGCGGGGGCCGGACGCCGAGGGCCTCTGGTTCGGCCGCGCGGGCGGCTCGG 22775

Qy   141 attcaaccgcctctccatcattgatattgcacactcccacccaaccactcgttggggacc 200
Db 22774 TCACCGGCGGCCTGTCCGTGATCGACCCCGAGCACGGCCGGCAGCCGATGACCGCGGAAC- 22714

Qy   201 tgcggatgaaccgaccgctacgcaatgatctcaacgtgagatctacaactacgttga 260
Db 22715 -----ACGAGGGCCGCGGCGGCGGCGATCATCTTCAGCCGCGAGATCTACAACTTCCGTGA 22661

Qy   261 gctgcgtaaagagctctcggattgggatataccttaatacttctggcgatggcgagcc 320
Db 22660 GCTGCGCGGCGAACTCACCTCGCACCGCCACCGGTTCAGGACGTCCTGCGATACCGAGGT 22601

Qy   321 aattgttgtcggtttccaccactgggcgagtccgtggtcgagcatctccgcggaatgtt 380
Db 22600 CGTCCTGCGCGGCGGCTACCTCCAGTGGGCGGCCGCGCGCTGGTCGAGAGGCTCAACGGGATGTT 22541

Qy   381 cggcattgccatttgggatacaaaggaaaagtcgctttccttgcggtgatcagttcgg 440
Db 22540 CGCCTTCGCCGTCTGGGACGAGCACAGCGAGGAACTCCTGCTGGTCGGCGACCGGATGGG 22481

Qy   441 catcaagccactgttctacgcaaccaccgagcatggcgaccgtgttcctccctcagagaagaa 500
Db 22480 CGTCAAGCCGCTCTACTACTGGCCGACCGCCCGCGGCTGCGCCTTCGCGTCCGGAGCCCAA 22421

Qy   501 gaccatcttggag 513
Db 22420 GGCCGTCCTGGCG 22408

RESULT 8
AAF77894
ID AAF77894 standard; DNA; 1872 BP.
XX
AC AAF77894;
XX
DT 04-JUN-2001 (first entry)
XX
DE Quorum sensing controlled gene qsc137 ORF.
XX
KW Quorum sensing; antibacterial; bacterial signalling;
KW opportunistic pathogen; immunocompromised; burn; cystic fibrosis;
KW immunosuppressive therapy; AIDS; ss.
XX
OS Pseudomonas aeruginosa.
XX

Db 22834 CATGGCCTGCTGCGGGGGCCGGACGCCGAGGGCCTCTGGTTCGGCCGCGCGGGCGGCTCGG 22775
Qy   141 attcaaccgcctctccatcattgatattgcacactcccacccaaccactcgttggggacc 200
Db 22774 TCACCGGCGGCCTGTCCGTGATCGACCCCGAGCACGGCCGGCAGCCGATGACCGCGGAAC- 22714
Qy   201 tgcggatgaaccgaccgctacgcaatgatctcaacgtgagatctacaactacgttga 260
Db 22715 -----ACGAGGGCCGCGGCGGCGGCGATCATCTTCAGCCGCGAGATCTACAACTTCCGTGA 22661
Qy   261 gctgcgtaaagagctctcggattgggatataccttaatacttctggcgatggcgagcc 320
Db 22660 GCTGCGCGGCGAACTCACCTCGCACCGCCACCGGTTCAGGACGTCCTGCGATACCGAGGT 22601
Qy   321 aattgttgtcggtttccaccactgggcgagtccgtggtcgagcatctccgcggaatgtt 380
Db 22600 CGTCCTGCGCGGCGGCTACCTCCAGTGGGCGGCCGCGCGCTGGTCGAGAGGCTCAACGGGATGTT 22541
Qy   381 cggcattgccatttgggatacaaaggaaaagtcgctttccttgcggtgatcagttcgg 440
Db 22540 CGCCTTCGCCGTCTGGGACGAGCACAGCGAGGAACTCCTGCTGGTCGGCGACCGGATGGG 22481
Qy   441 catcaagccactgttctacgcaaccaccgagcatggcgaccgtgttcctccctcagagaagaa 500
Db 22480 CGTCAAGCCGCTCTACTACTGGCCGACCGCCCGCGGCTGCGCCTTCGCGTCCGGAGCCCAA 22421
Qy   501 gaccatcttggag 513
Db 22420 GGCCGTCCTGGCG 22408

RESULT 7
AAX05110/c
ID AAX05110 standard; DNA; 30001 BP.
XX
AC AAX05110;
XX
DT 07-APR-1999 (first entry)
XX
DE S. aureofaciens DNA in cosmid clones LP2-127 and LP2-128.
XX
KW Chlortetracycline; tetracycline; biosynthetic; actinomycete;
KW S. ariseofuscus; S. ambofaciens; antibiotic resistance gene;
KW S. lividans; S. aureofaciens; cosmid clone; antibiotic; ss.
XX
OS Streptomyces aureofaciens.
XX
PN US5866410-A.
XX
PD 02-FEB-1999.
XX
PF 07-JUN-1995; 95US-0474933.
XX
PR 22-SEP-1993; 93US-0125468.
PR 26-JUL-1990; 90US-0558039.
PR 26-JUL-1990; 90US-0558040.
PR 15-JAN-1992; 92US-0821109.
PR 15-JAN-1992; 92US-0821419.
PR 07-JUN-1995; 95US-0474933.
XX
PA (AMCY) AMERICAN CYANAMID CO.
XX
PI Fantini SE, Lotvin JA, Ryan MJ, Strathy N;
XX
DR WPI; 1999-141936/12.
XX
PT Two plasmids for cloning the biosynthetic pathways of tetracycline,
PT chlortetracycline, and their derivatives - comprise an origin of
PT replication, cohesive end sites and optionally, an antibiotic
PT resistance gene
XX
PS Disclosure; Fig 4A-L; 39pp; English.
XX

XX
PN   WO200118248-A2.
XX
PD   15-MAR-2001.
XX
PF   01-SEP-2000; 2000WO-US24141.
XX
PR   03-SEP-1999; 99US-0153022.
XX
PA   (IOWA) UNIV IOWA RES FOUND.
PA   (QUOR-) QUORUM SCI INC.
XX
PI   Whiteley M, Lee KM, Greenberg EP, Muh U;
XX
DR   WPI; 2001-265973/27.
XX
PT   Identifying modulators of quorum sensing signaling in Pseudomonas
PT   aeruginosa bacteria, useful for treating infections in
PT   immunocompromised patients -
XX
PS   Claim 46; Page 94; 115pp; English.
XX
CC   Bacteria signal to one another to coordinate expression of specific genes
CC   in a cell density dependent fashion. This "bacterial signalling" is
CC   called "quorum sensing and response". Quorum sensing allows a bacterial
CC   species to sense its own number and regulate gene expression according to
CC   population density. The present sequence is an open reading frame (ORF)
CC   of a Pseudomonas aeruginosa quorum sensing controlled gene. Inhibitors of
CC   quorum sensing signaling renders a bacterial population more susceptible
CC   to treatment. The present invention relates to a method for identifying
CC   modulators of quorum sensing signaling in Pseudomonas aeruginosa
CC   bacteria. Modulators of quorum sensing signaling may be used to treat P.
CC   aeruginosa infections. P. aeruginosa is an opportunistic pathogen of
CC   immunocompromised individuals (burn patients, cystic fibrosis patients,
CC   patients undergoing immunosuppressive therapy and patients with AIDS).
XX
SQ   Sequence 1872 BP; 302 A; 626 C; 615 G; 329 T; 0 other;

Query Match   4.0%;   Score 77.4; DB 22;   Length 1872;
Best Local Similarity   50.3%;   Pred. No. 2.8e-12;
Matches 217;   Conservative   0;   Mismatches 211;   Indels   3;   Gaps   1;

Qy   91 cgtggtcctcgacgatgccggcactggcgccgatgacgccgatgcagcgtttggattcaaccgc 150
          || |||||| ||| || ||||| |||||||| ||||| || ||||| | | |||||| |
Db   130 cgcgggccggatgccgagggcatcggaacgcacggcaaagccctgctgggtcaccggcgg 189

Qy   151 ctctccatcattgatattgcacactcccaaccactgcgttggggacctgcggatgaa 210
          || |||||| || || || | | | | || |||| || |||||| || ||
Db   190 ctggcggtcatcgacctcagcgggcgtgcagccgtgtccatcgctttcccaccggc 249

Qy   211 ccgaccgctacgcaatgactttcaacggtgagatcctacaactacgttgagctcgtaaa 270
          || || | || | | ||| |||||| ||||| || |||| || || |||| |
Db   250 cagga--ggtcaccctcgctcacacggcgaggtgtacaaccacgatgcctgcgag 306

Qy   271 gagctctcggatttggatatacctttaatacttctggcgatgcgagccaattgttgtc 330
          | | | ||| | |||| | |||| | |||||||| || ||||| ||| |||
Db   307 cggttgcgcggccggacatgagttccgcaccgggataccgcaggtggtcctgcac 366

Qy   331 ggtttccaccactggggcgagtccggtcgtcgagcatctccgcggaatgttcggcattgcc 390
          |||| || ||| || || || ||| | | ||| | |||| |||| ||| ||||||||
Db   367 gcctatctgcaatggggcgagcggttgttgcagtacctgagtggatgttcgccttcgcc 426

Qy   391 atttggatacaaaggaaaagtcgctttcttcctggcgtgatcagttcggcatcaagcca 450
          | ||||| | | | | |||| || || || ||| ||| | ||||||||| ||
Db   427 gtcttcgatggccgcgacggccacctgctcggtcgcgacggccctgggcgcatcaagccg 486

Qy   451 ctgttctacgcaccagcgagcatgcaccgtgttctcctcagagaagagaccatcttg 510
          ||||| || ||||||| || ||| || || || || | ||| || |||| |
Db   487 ctgtattacgcgcggcgcaccgcgagggactgctgttcggctcggagatcaagtcatcctg 546

Qy   511 gagatggccga 521
          || || |||||
Db   547 gcgcatccgga 557

RESULT 9
AAH41225/c
ID   AAH41225 standard; DNA; 349980 BP.
XX
AC   AAH41225;
XX
DT   29-OCT-2001 (first entry)
XX
DE   Pyrococcus abyssi genomic fragment #4.
XX
KW   Hyperthermophilic archaeon; hyperthermophilic protein; ds.
XX
OS   Pyrococcus abyssi.
XX
FH   Key   Location/Qualifiers
FT   misc_feature   1..49980
FT   /*tag= a
FT   /note= "This sequence overlaps with the 3' end of
FT   AAH41224"
FT   misc_feature   300001..349980
FT   /*tag= b
FT   /note= "This sequence overlaps with the 5'end of
FT   AAH41226"
XX
PN   FR2792651-A1.
XX
PD   27-OCT-2000.
XX
PF   21-APR-1999; 99FR-0005034.
XX
PR   21-APR-1999; 99FR-0005034.
XX
PA   (CNRS) CNRS CENT NAT RECH SCI.
PA   (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
PI   Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
PI   Querellou J, Weissenbach J, Saurin W, Heilig R;
XX
DR   WPI; 2001-126236/14.
XX
PT   New nucleotide sequences isolated from Pyrococcus abyssi encode
PT   proteins useful in industry -
XX
PS   Claim 1; Page 429-524; 1657pp; French.
XX
CC   The present invention relates to the genomic sequence of Pyrococcus
CC   abyssi and P. abyssi proteins (see AAB96053-AAB96842). P. abyssi is a
CC   hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal
CC   vents. The present sequence is a fragment of the genomic sequence of P.
CC   abyssi. The 5' end of this sequence overlaps with the 3' end of AAH41224
CC   and the 3' end of this sequence overlaps with the 5' end of AAH41226. The
CC   proteins of the present invention have various potential industrial uses,
CC   since the proteins are stable at very high temperatures, some up to 110
CC   degrees centigrade.
CC   Note: This patent is in the same patent family as WO200065062, which
CC   contains additional sequences as shown in AAB99132-AAB99143,
CC   AAH75903-AAH75920 and AAG66436.
XX
SQ   Sequence 349980 BP; 99421 A; 76616 C; 77444 G; 96499 T; 0 other;

Query Match   3.9%;   Score 74.6; DB 22;   Length 349980;
Best Local Similarity   50.18%;   Pred. No. 2.9e-10;
Matches 213;   Conservative   0;   Mismatches 209;   Indels   3;   Gaps   1;

Qy   82 atgcgccaccgtggtcctgacgatgccgacactggcatgccgatgacgccgatgcagcgtttgga 141
           |||| ||||||||||||| | ||||| |||||| || |||||| || |||||| || |||| ||
Db 209411 ATTCGGCATCGTGGACCTGAGCATGCATCGTGGAGTTTATGTTGATGATAATGTTAGCCTTGGT 209352

Qy   142 ttcaaccgcctctccatcattgatattgcactcccaccacctgcgttggggacct 201
           || | ||| | |||||| || ||| || | | | | || |||| ||

```
Db 209351 CACGTTAGGCTTGCCATCATTGACCTCTCAC---CCAAGGGGCACCAGCCTATGAAGTAC 209295
              |||  ||||||||| ||||||| ||  ||||        |||  ||||||||||||||||
Qy    202 gcggatgaaccgacccgctacgcaatgactcttcaacgtgagatctacaactacgttgag 261
           ||  ||  ||| ||| |||    | || |||    |||| || ||  ||| ||||| ||
Db 209294 GAGAAAGATGGTAAAGAGAGTTTCGATCGTTTACAATGGGGAAATTTATAACTTTATGGAA 209235
          ||| ||    ||| ||||| || ||  ||||| |  | || ||||| ||| ||| ||
Qy    262 ctgcgtaaagagctctcggattggatataccttaatactctcggcgatggcgagcca 321
           || ||  |||||  |||  || ||||| ||  ||   |||  || ||  ||||||
Db 209234 ATTCGAAAAGAGAGCTTGAGGAGAGAGGGTTACACTTTCACCTCTAACACTGACACCGAGGTT 209175
          || |   |||||  ||  ||||| || || ||| || | ||||  || || ||||| |||
Qy    322 atgtgttcggttccaccactggggcgagtcgtggtcgagcatctccggcgaatgttc 381
           || ||  ||||||    | ||||| | ||||| || || ||| ||  || ||||| |
Db 209174 ATCCTTGCCGCCTACCTAGAGTTGGGGCTTTGATTGTGTGGAGAGGTTCAATGGCATGTGG 209115
          | ||  || || |||| || || || || |  |||| || |||||  ||| |||| |||
Qy    382 ggcattgccatttgggatacaaagaaaagtcgctttcctttcctgcgctgatcagttcggc 441
           | ||||| ||  ||| ||  ||| |||| |  |||    |   ||| |||| |||  ||
Db 209114 GCTTTCGTGATCTATGCAAGCTCATCGTAAGAACATCCTGTTTCTAAGCAGGACAGATTCCGA 209055
          | || || ||| || | ||| | | || |||| || |  ||  ||||| ||| ||| ||||
Qy    442 atcaagccactgttctacgcaaccaccgagcactgttctcctcagaagaaag 501
           |  |||||| || ||  || |||| ||| |  || ||||  ||||||||||
Db 209054 ATAAAGCCCCTGTATTATTACTATGACGGCAAAAACATAATCTTCAGCTCAGAAATCAAG 208995
          || |||||| || ||   ||   |  ||  |||  |||   |   ||||||||||||||
Qy    502 accat 506
           |||
Db 208994 GGAAT 208990
```

RESULT 10

AAA58471

ID   AAA58471 standard; DNA; 58857 BP.
XX
AC   AAA58471;
XX
DT   31-OCT-2000 (first entry)
XX
DE   Nucleotide sequence of the bleomycin (BLM) gene cluster ORFs 8-30.
XX
KW   BLM gene cluster; bleomycin gene cluster; polyketide metabolite;
KW   bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine;
KW   thiazoline; bithiazoline; microbial metabolite; sugar; ss.
XX
OS   Streptomyces verticillus.
XX
FH   Key   Location/Qualifiers
FT   CDS   223..564
FT   /tag= a
FT   /note= "ORF 30; encodes AAB07556"
FT   CDS   561..2309
FT   /transl_except= (pos: 1..3, aa: Met)
FT   /note= "ORF 29; encodes AAB07557"
FT   /tag= b
FT   CDS   2767..3486
FT   /tag= c
FT   /note= "ORF 28; encodes AAB07558"
FT   CDS   3527..5593
FT   /tag= d
FT   /note= "ORF 27; encodes AAB07559"
FT   CDS   5806..12294
FT   /transl_except= (pos: 1..3, aa: Met)
FT   /note= "ORF 26; encodes AAB07560"
FT   /tag= e
FT   CDS   12291..15491
FT   /tag= f
FT   /note= "ORF 25; encodes AAB07561"
FT   CDS   15488..21013
FT   /note= "ORF 24; encodes AAB07562"
FT   /tag= g
FT   CDS   21010..24666
FT   /tag= h
FT   /transl_except= (pos: 1..3, aa: Met)
FT   /note= "ORF 23; encodes AAB07563"
FT   CDS   24663..32690
FT   /tag= i
FT   /note= "ORF 22; encodes AAB07564"
FT   CDS   32893..34830
FT   /tag= j
FT   /note= "ORF 21; encodes AAB07565"
FT   CDS   34827..35804
FT   /tag= k
FT   /transl_except= (pos: 1..3, aa: Met)
FT   /note= "ORF 20; encodes AAB07566"
FT   CDS   35818..37302
FT   /tag= l
FT   /transl_except= (pos: 1..3, aa: Met)
FT   /note= "ORF 19; encodes AAB07567"
FT   CDS   37299..39215
FT   /tag= m
FT   /transl_except= (pos: 1..3, aa: Met)
FT   /note= "ORF 18; encodes AAB07568"
FT   CDS   39301..47181
FT   /tag= n
FT   /note= "ORF 17; encodes AAB07569"
FT   CDS   47178..49985
FT   /tag= o
FT   /transl_except= (pos: 1..3, aa: Met)
FT   /note= "ORF 16; encodes AAB07570"
FT   CDS   49982..51001
FT   /tag= p
FT   /transl_except= (pos: 1..3, aa: Met)
FT   /note= "ORF 15; encodes AAB07571"
FT   CDS   50998..52386
FT   /tag= q
FT   /transl_except= (pos: 1..3, aa: Met)
FT   /note= "ORF 14; encodes AAB07572"
FT   CDS   52383..52946
FT   /tag= r
FT   /note= "ORF 13; encodes AAB07573"
FT   CDS   53018..54190
FT   /tag= s
FT   /note= "ORF 12; encodes AAB07574"
FT   CDS   54187..55824
FT   /tag= t
FT   /note= "ORF 11; encodes AAB07575"
FT   CDS   55821..56093
FT   /tag= u
FT   /transl_except= (pos: 1..3, aa: Met)
FT   /note= "ORF 10; encodes AAB07576"
FT   CDS   56090..57586
FT   /tag= v
FT   /transl_except= (pos: 1..3, aa: Met)
FT   /note= "ORF 9; encodes AAB07577"
FT   CDS   57583..58857
FT   /tag= w
FT   /note= "ORF 8; encodes AAB07578"
XX
PN   WO200040704-A1.
XX
PD   13-JUL-2000.
XX
PF   06-JAN-2000; 2000WO-US00445.
XX
PR   06-JAN-1999; 99US-0115435.
PR   05-FEB-1999; 99US-0118848.
PR   05-JAN-2000; 2000US-0477962.
XX
PA   (REGC) UNIV CALIFORNIA.
XX
PI   Shen B, Du L, Sanchez C, Chen M, Edwards DJ;
XX
DR   WPI; 2000-465974/40.
DR   P-PSDB; AAB07556, AAB07557, AAB07558, AAB07559, AAB07560, AAB07561,
DR   AAB07562, AAB07563, AAB07564, AAB07565, AAB07566, AAB07567, AAB07568,
DR   AAB07569, AAB07570, AAB07571, AAB07572, AAB07573, AAB07574, AAB07575,
DR   AAB07576, AAB07577, AAB07578.

```
XX
PT  New bleomycin gene cluster components useful for peptide and/or
PT  polyketide metabolites, especially bleomycin, production and for
PT  chemically modifying biological molecules -
XX
PS  Claim 8; Page 97-136; 162pp; English.
XX
CC  The present sequence represents the BLM (Bleomycin) gene cluster,
CC  containing open reading frames (ORFs) 8-30. The proteins encoded
CC  by the gene cluster are useful for producing peptides and/or polyketide
CC  metabolites, especially bleomycin or bleomycin analogues. They are
CC  also useful for chemically modifying biological molecules to produce
CC  branched methyl groups, and for coupling amino acids and fatty
CC  acids. They may be reacted with an apo-carrier protein and coenzyme A
CC  to produce a holo-carrier protein. The BLM gene cluster or catalytic
CC  domains can be used individually or collectively to produce
CC  thiazolidine, thiazoline, bithiazoline and bithiazoline-containing
CC  microbial metabolites. The BLM gene cluster may also be used to produce
CC  sugars.
XX
SQ  Sequence 58857 BP; 7256 A; 25139 C; 19353 G; 7109 T; 0 other;

Query Match         3.5%;  Score 67.8;  DB 21;  Length 58857;
Best Local Similarity   42.9%;  Pred. No. 1.2e-08;
Matches  748;  Conservative  0;  Mismatches  967;  Indels  30;  Gaps  7;
```

```
Qy      57  tgcactcgagcggggccttgccatgcgatgcgccaccgtgtcctgacgatgccggcacttg   116
            || |||  ||  |  ||  || | ||||||| || ||||  |  | || |||| |
Db   37358  tgaactcaccgacgccgatggcgccggaaggcacctg   37417

Qy     117  gcatgacgacgccgatgcagcgtttgattcaaccgcctctccatcattgatattgacactc   176
            ||  | ||  ||  | | ||| ||||| | ||||  |  | || ||||| |
Db   37418  ggtctcgcccaccgcggcgcgggccgacacgccgctgatcgccgtgatcgccccga   37477

Qy     177  ccaccaacactgcgttggggacctgcggatgaaccgacccgaccgatcacgcaatgactttcaa   236
            ||  || |||  |  | || |  |||| || |||   |  | ||||
Db   37478  cgccgacgcagccgtcgccggaccgcacgcaccgccggctcg------tcgtcaa   37531

Qy     237  cggtgagatcacactacgttgagctgctggctaaagagctctggatttgggatattacctt   296
            |||| || ||  |  || |||  | | |||  |||| |  ||||||  ||| ||
Db   37532  cggcgagttctacggtaccggggagtccggcggactgcggcggctggtt   37591

Qy     297  taatacttctggcgatggcgagccaattgttgtcggttccaccactgggcggggctccgt   356
            || || | ||  | ||| ||  ||  | ||  | || |   |  || ||| |  || ||
Db   37592  ccggacccgcagcgacagcgagatcgccctccacctgcgggacgcgcggggc   37651

Qy     357  ggtcggacatctccgggaatgttccgcattgccatttggatacaaaaggaaaagtcgct   416
            ||| || | |||||| ||  |  | || | ||  | ||  | || ||  |  |||
Db   37652  actggagcggctgcggcgcagttcgctccttcgtcctcgcctgggacgaacgccgccaccct   37711

Qy     417  tttccttgcgtgatcagttcgcatcaagccactgttctacgccaacaccgagcatgg   476
            || || | ||  | |||| ||||| || || |  | | ||| |||| | ||  ||
Db   37712  cttcgccgccgaccgttcggcggcgtcagccgtcaaaccctacta-----caccgagcgcga   37765

Qy     477  caccgtgtttcctccagagaagaagaccatcttggagatggccgaggagatgaatctaga   536
            |||| | | |   ||||| ||||| | || || || | |   |  ||  |  || |||
Db   37766  cggcggctcacgctacggcgcgggccctcgtctccgtcgcggcgccgcgggcgccgcccg   37825

Qy     537  tctgggccttgataagcgccgcaccattgagcactactggacactgacagttcgcc   596
            | ||| ||  | |||| |||||| || |  | ||||  |||| || | |  |||
Db   37826  ctgggacaccggcgccttcgcgcgcgcaccgcagctgccaccatcgccgcgccacccct   37885

Qy     597  agatacccttcacgcgcagatttcccgcttggagtcaggctcgaccacggcaacagttcgtcc   656
            ||||| || || ||| || | | |  |  || || || | | ||  ||| ||| || ||
Db   37886  cttcgcacctcgatggcgacgacgacaccccctcgacgcgccacgggcac   37945

Qy     657  gggcgcaagctggaacagacagagcgttacttcaagcctcagttcccagtacagaaggtcgt   716
            ||| | ||| || || ||| | || | |  | || |||| | | | ||| || |  |||
Db   37946  ccgcgtcacccactgggacctcgactacccgcc------cgccggcgaactccgcgc   37999

Qy     717  aaaggtaaggagcaggagaccctctcgatccagttcccagtggttgggatatagcgtcga   776
            |||| | || || |  |  || | ||| |  | || || || || ||| || |  |
Db   38000  ccgcgtcacccactgggacctcgactacccg
```

```
Db   38000  ccggggagcgtcggacgaccaccctggacgacgcggtacgcggacgacgagccgtacg   38059
Qy     777  aaagcatatgcgtgcgacgtgaccgctaggcctgttccttccggcgcggcattgcac   836
            || |   |   ||  | | || ||||  ||  ||||||||  || ||||  |   ||
Db   38060  gttgcgtacgccgcgacgtgccccctcgcctgccacctcagcgcggcggcctggactcctc   38119
Qy     837  cgcaattgcgccgtcgaaagcgccacaacccctgaccgctcacctcaccacccggctt   896
            || ||  |||| ||| || ||||  |   || |||| || |||| |||| |||||| ||
Db   38120  cgccgtcgccgcctcgccgccgccacccc---ggctcaccgccttcaccgtccgcctt   38176
Qy     897  cgagcggaagctactcggagtcgatggctgcggagtcgccgccgcgcgatggcgc   956
            |||| | ||| ||  ||| ||||| ||  | ||  || ||| || |||| |||| |
Db   38177  cgacgacccgcgccttcgacgagagcggcgtcgccgagagcgacccacttgcgggacgtcgt   38236
Qy     957  tgagcacctcgtgaagatcgtctcgctgaggaatacgccacgcgattcctaagatcat   1016
            || |||||| || || || || ||| | ||  |||  |   || | || |||| ||
Db   38237  cgaccaccgcgaagtcgctcggaacgcgccacttcgcggaccacctgcgggacgtcgt   38296
Qy    1017  gtggtacttggatgatcctgtagctgactgacggagaactgcacgcatcgtccgctgcagc   1076
            ||||||||| || |||  | | | | ||| || |||| ||||  ||  ||  |  |||||
Db   38297  ccgcgccggcgagatggtgcaggaagaactgcacgcatcgcacgccggtacctgcacagcgc   38356
Qy    1077  ggaagcacgtaagcacg---tcaaggttgtgctgtctggcggaggcgccagatgagctgtt   1133
            |||||||| ||   | |   ||||| |||  |  |  | ||  |  |   || | |  ||
Db   38357  gcatcaagaaggcgggattcaccgcgtcccgcggggaggcggggacgaactgtt   38416
Qy    1134  cggtggatacaccatttacaaagagccgctatcgttgctcccattggagaaatacctt--   1191
            |  |||| |  ||| || | ||| | || | |||  | | ||  |||| || | ||
Db   38417  cctcggctaccccagttccgcaaggacctgagcgacgtccccgacggcatgaag   38476
Qy    1192  ---tccccactagtaaagcctggaaagctcagcaagttctcgacggccactg   1248
               ||  | ||  |  || ||| ||  | | || ||  | ||| |||| | |
Db   38477  caaggcgaccggcgtacgccggtcgtgacgccgggtcctcgccgccgctgcg   38536
Qy    1249  ggcaagtccctcttgagcgtggctgctgccatgaccatggaagagagcgctactacggcaacgct   1308
            ||| ||| ||| | || ||||| | | |||  |  || |||| ||||  | || || ||  |||
Db   38537  caccctctcggcacccgaccctccggcttcctccgatcgtcgacgcgccaccctggcgt   38596
Qy    1309  cgctcctcaattcgagcagatgacagcaacgcgtattccatgggcaaagcgaatgggac   1368
            || | || | ||||| |||| | |  || ||||  | |||  |  |  | |  ||||||
Db   38597  caccagccgctaccgtcgcgccgacttgcgcgccgaacttgatccgtagcccgc   38656
Qy    1369  caccgaagtcactgacgatcacatccgcacgccgcgcgccgccgccaacttgatccagtagcccgc   1428
            ||||| || | |  ||| | | |||||| |  ||  || ||||| |  ||  |   | ||||
Db   38657  cgccgcgccccctgccctcgcgccgcggctgcggcgcgggccgccggggcga   38716
Qy    1429  atgcaaccctgatctgttcacctggatcgtcggacatccggtcaaggctgacaa   1487
            || ||  || | |  | || |  || | || ||   |  |  | ||| |  |||
Db   38717  ccaggccacctaccctcttcgccaagaccctggctcgtctcgccgccgaacg   38776
Qy    1488  gatcaacatggcgaactccctggctgcgagtccattcttggataaggaagtttcaa   1547
            || || |||| |||| ||||| || |  || ||| | |||   | ||   | ||
Db   38777  cctcgacgcggccaggccgtcgaggtcgaggctcgcccctccctcgaccacctcttcga   38836
Qy    1548  ggttgcagagaccattcctacgatctgaagattgcaacgtaccaccaagtacgcgct   1607
            |||| |||||  |||  ||| | |  || |  ||  |  || || ||| | |||||
Db   38837  cctcgtccgcacggccactgttccgccctgcgcccgccgagtgaccgaagtccaagtacccggct   38896
Qy    1608  gcgcagggcactcgacgagatgttccgcctcacgtttgcaccgcaagaagctgggctt   1667
            || ||| || || || | | | || || ||  ||||| ||||  ||| ||  ||| ||
Db   38897  gcgggccgccgatcgccgccatgggcgctgccgcgagtgaccgcaaacaggcctt   38956
Qy    1668  cctgttccatgcgcagctacagctccggctgccggcgatgagctgtttcggttgggcgcaggacac   1727
            || ||||||  |||  ||||  |||  |||| |||| ||||| ||  ||||  |  |||| ||
Db   38957  cctcgcacctcgatggcgacgacgacaccccctcgacgcgccagaacttgcgctgtggatgctgaa   39016
Qy    1728  cattaaggaatccggtactgaagatatcttcaacagaggtatatgctgaa   1787
            ||  |||| | ||| || |||| |||| | ||||  | || | |||||
Db   39017  cggacggcggcggggacgacgaccccctcttcgacccgacccgacgacgcgctgtcgctgga   39076
Qy    1788  cgagc   1792
            |||||
Db   39077  ccggc   39081
```

```
RESULT 11
AAC78035
ID   AAC78035 standard; cDNA; 1470 BP.
XX
AC   AAC78035;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human cancer associated gene sequence SEQ ID NO:429.
XX
KW   Human; cancer associated gene; cancer antigen; detection; cancer;
KW   diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;
KW   antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW   antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW   dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW   vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW   immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW   allergic reaction; graft versus host disease; organ rejection;
KW   haemostatic; thrombolytic; cardiovascular disorder; infection;
KW   neurological disease; drug screening; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055350-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05882.
XX
PR   12-MAR-1999;  99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587533/55.
DR   P-PSDB; AAB43826.
XX
PT   Novel isolated nucleic acids comprising sequences encoding peptides
PT   useful for treating or diagnosing e.g. cancer -
XX
PS   Claim 1; Page 970; 2352pp; English.
XX
CC   AAC77607 to AAC78448 encode the human cancer associated proteins given
CC   in AAB43398 to AAB44239. The proteins can have activities based on the
CC   tissues and cells the genes are expressed in. Example of activities
CC   include: cytostatic; proliferative; vulnerary; immunomodulator;
CC   antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC   antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC   dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC   nootropic; vasotropic; antipsoriatic; antiangiogenic and antiangiogenic. The
CC   polynucleotides and polypeptides can be used for preventing, treating or
CC   ameliorating medical conditions and diagnosing pathological conditions.
CC   Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC   the present invention may be used to treat immune disorders by activating
CC   or inhibiting the proliferation, differentiation or mobilisation of
CC   immune cells, to treat disorders of haematopoietic cells, autoimmune
CC   disorders, allergic reactions, graft versus host disease and organ
CC   rejection, modulate haemostatic or thrombolytic activity, modulate
CC   inflammation, cancers, cardiovascular disorders, neurological disease and
CC   bacterial or viral infections. The peptides, nucleotides, antibodies,
CC   agonists and antagonists may be also be used in drug screens. AAC78449 to
CC   AAC78457 and AAB44240 represent sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 1470 BP; 336 A; 380 C; 470 G; 277 T; 7 other;

Query Match       2.48%;  Score 45.6;  DB 21;  Length 1470;
Best Local Similarity  49.2%;  Pred. No. 0.007;
Matches 120;  Conservative 0;  Mismatches 124;  Indels 0;  Gaps 0;
```

```
Qy          1409 acttgatccagtagcccgatgcaacacctggatctgtcacctggatgcgcggcgaca 1468
                 ||||||||  ||  ||| ||| |||||| |||||||||| || ||||| ||| |||
Db           451 acttggccgtctggtcaacctccagcacccgatcctccaacaacaagctggtcacct 510

Qy          1469 tcctggtcaaggctgacaagatcaacatggcgaactccctggacgccgagttccattct 1528
                 || ||||  ||| |||||||| ||| ||| |||||| ||||| ||||| ||||||| ||
Db           511 tgcctgtcagcttgctgctcagctcaagaacctgaagtggtggacctgaagataacccc 570

Qy          1529 tggataaggagtttcaaggtgcagagaccattcctcacatctgaagattgccaacg 1588
                 ||||| |  ||||||  |||| ||| || |||| ||| || ||| ||| |||  ||
Db           571 tggatcctgctcctggccaagtggcaggtgactgcttgctggtgagaagcagtgtaagcagt 630

Qy          1589 gtaccaccagtacgctgcgcgcagggcactcgacagattgttccgcctcacgtttgc 1648
                 ||| ||||  ||||| || ||   |||| | || || ||| || ||| ||||| ||
Db           631 gtgcaacaaggtgttacagcacatgaaggccgtgcaggcagatcaggaggcgggagaggc 690

Qy          1649 accg 1652
                 ||||
Db           691 agcg 694
```

```
RESULT 12
AAS64941
ID   AAS64941 standard; cDNA; 1518 BP.
XX
AC   AAS64941;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #745.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,   Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG00754.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 745; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
```

Tue Sep 24 09:40:54 2002

us-09-786-474-1.rng   Page 14

```
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note:The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1518 BP; 367 A; 409 C; 480 G; 262 T; 0 other;

Query Match             2.4%;  Score 45.6;  DB 23;  Length 1518;
Best Local Similarity   49.2%;  Pred. NO. 0.0071;
Matches  120;  Conservative   0;  Mismatches  124;  Indels   0;  Gaps   0;

Qy       1409 actttgatccagtagccgcccgcatgcaacacctggatctgttcacctggatgcgcggcgaca 1468
                |||||| ||  | ||  ||||||||||||||| ||| |  ||| |  |
Db        833 actttggccgtctggtcaacctccagcacctggatctcctcaacaacaagctggtcacct 892

Qy       1469 tcctgtcaaggctgacaagatcaacatggcgaactccctgagctgcgagttccattct 1528
                | |||  |||  ||  | |||||  | ||||| || || ||
Db        893 tgcctgtcagctttgctcagctcaagaacctgaagtggtggacctgaaggataaccccc 952

Qy       1529 tggataaggaagtttcaaggttgcagagaccattcctacgatctgaagattgccaacg 1588
                ||||| | ||||  |||| ||||  |  |  |||  | ||  |||
Db        953 tggatcctgtcctggccaaggtggcaggtgactgcttggatgagaagcagtgtaagcagt 1012

Qy       1589 gtaccaccaagtacgcgctgcgcaggcactcgagcagattgttccgcctcacgttttgc 1648
                ||||  || |  ||| |||||  || ||| |
Db       1013 gtgcaaacaaggtgttacagcacatgaaggccgtgcaggcagatcaggagcggggagagc 1072

Qy       1649 accg 1652
                | ||
Db       1073 agcg 1076

RESULT 13
AAS22442
ID   AAS22442 standard; cDNA; 3331 BP.
XX
AC   AAS22442;
XX
DT   24-OCT-2001 (first entry)
XX
DE   Human cDNA encoding a novel human protein #8.
XX
KW   Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;
KW   immunomodulatory; cytostatic; neuroprotective; vulnerary; nootropic;
KW   anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW   antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW   thrombolytic; immunogen; antibody; neurological disorder;
KW   Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW   tissue regeneration; immune disorder.
XX
OS   Homo sapiens.
XX
PN   WO20015437-A2.
XX
PD   02-AUG-2001.
XX
PF   25-JAN-2001; 2001WO-US02623.
XX
PR   25-JAN-2000; 2000US-0491404.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT;
XX
DR   WPI; 2001-451939/48.
DR   P-PSDB; AU14137.
XX
PT   Isolated polypeptides useful for treating anti-inflammatory diseases,
```

```
PT   nervous system disorders, and for regenerating bone and cartilage -
XX
PS   Claim 1: Page 181-183; 894pp; English.
XX
CC   The invention relates to polynucleotides encoding novel human
CC   proteins or their active domains. The polypeptides, polynucleotides and
CC   antibodies raised against the polypeptides are used in a method of
CC   treatment of a mammal and prevention of disorders caused by the aberrant
CC   protein expression or activity. The polypeptides can be used as
CC   molecular weight markers, food supplements, and in antibody production.
CC   The polypeptides are used to identify compounds which bind to the
CC   polypeptides. Polynucleotides of the invention are used as probes and
CC   primers, for sequencing, for chromosome or gene mapping, in the
CC   production of recombinant proteins, and in generating anti-sense DNA or
CC   RNA and in gene therapy.  Polypeptides of the invention can be used to
CC   target drugs to a tumour, in assays to determine biological activity, to
CC   raise antibodies/elicit an immune response, to determine quantitative
CC   protein levels, as tissue markers, and to isolate receptors or ligands.
CC   Polypeptides of the invention may also be useful in treating platelet
CC   disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC   ligament and/or nerve tissue, wound healing, treating burns, promoting
CC   the proliferation, differentiation and survival of stem cells, as a
CC   contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC   Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
CC   sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC   fungal infection or from autoimmunity, cancer, allergy, asthma,
CC   graft-versus-host disease, eczema, haemophilia, thrombosis,
CC   anti-inflammatory diseases, nervous system disorders, and infection.
CC   The present sequence encodes a protein of the invention.
XX
SQ   Sequence 3331 BP; 872 A; 824 C; 879 G; 756 T; 0 other;

Query Match             2.4%;  Score 45.6;  DB 22;  Length 3331;
Best Local Similarity   49.2%;  Pred. No. 0.011;
Matches  120;  Conservative   0;  Mismatches  124;  Indels   0;  Gaps   0;

Qy       1409 actttgatccagtagccgcccgcatgcaacacctggatctgttcacctggatgcgcggcgaca 1468
                |||||| ||  | ||  ||||||||||||||| ||| |  ||| |  |
Db        868 actttggccgtctggtcaacctccagcacctggatctcctcaacaacaagctggtcacct 927

Qy       1469 tcctgtcaaggctgacaagatcaacatggcgaactccctgagctgcgagttccattct 1528
                | |||  |||  ||  | |||||  | ||||| || || ||
Db        928 tgcctgtcagctttgctcagctcaagaacctgaagtggtggacctgaaggataacccc 987

Qy       1529 tggataaggaagtttcaaggttgcagagaccattcctacgatctgaagattgccaacg 1588
                ||||| | ||||  |||| ||||  |  |  |||  | ||  |||
Db        988 tggatcctgtcctggccaaggtggcaggtgactgcttggatgagaagcagtgtaagcagt 1047

Qy       1589 gtaccaccaagtacgcgctgcgcaggcactcgagcagattgttccgcctcacgtttgc 1648
                ||||  || |  ||| |||||  || ||| |
Db       1048 gtgcaaacaaggtgttacagcacatgaaggccgtgcaggcagatcaggagcggggagagc 1107

Qy       1649 accg 1652
                | ||
Db       1108 agcg 1111

RESULT 14
AAS22678
ID   AAS22678 standard; cDNA; 3690 BP.
XX
AC   AAS22678;
XX
DT   24-OCT-2001 (first entry)
XX
DE   Human cDNA encoding a novel human protein #244.
XX
KW   Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;
KW   immunomodulatory; cytostatic; neuroprotective; vulnerary; nootropic;
KW   anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW   antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW   thrombolytic; immunogen; antibody; neurological disorder;
```

```
KW   Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW   tissue regeneration; immune disorder.
XX
OS   Homo sapiens.
XX
PN   WO200155437-A2.
XX
PD   02-AUG-2001.
XX
PF   25-JAN-2001; 2001WO-US02623.
XX
PR   25-JAN-2000; 2000US-0491404.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-451939/48.
DR   P-PSDB; AAU14373.
XX
PT   Isolated polypeptides useful for treating anti-inflammatory diseases,
PT   nervous system disorders, and for regenerating bone and cartilage -
XX
PS   Claim 1; Page 673-674; 894pp; English.
XX
CC   The invention relates to polynucleotides encoding novel human
CC   proteins or their active domains. The polypeptides, polynucleotides and
CC   antibodies raised against the polypeptides are used in a method of
CC   treatment of a mammal and prevention of disorders caused by the aberrant
CC   protein expression or activity. The polypeptides can be used as
CC   molecular weight markers, food supplements, and in antibody production.
CC   The polypeptides are used to identify compounds which bind to the
CC   polypeptides. Polynucleotides of the invention are used as probes and
CC   primers, for sequencing, for chromosome or gene mapping, in the
CC   production of recombinant proteins, and in generating anti-sense DNA or
CC   RNA and in gene therapy. Polypeptides of the invention can be used to
CC   target drugs to a tumour, in assays to determine biological activity, to
CC   raise antibodies/elicit an immune response, to determine quantitative
CC   protein levels, as tissue markers, and to isolate receptors or ligands.
CC   Polypeptides of the invention may also be useful in treating platelet
CC   disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC   ligament and/or nerve tissue, wound healing, treating burns, promoting
CC   the proliferation, differentiation and survival of stem cells, as a
CC   contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC   Alzheimer's, Parkinson's and Huntington's diseases, amylotrophic lateral
CC   sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC   fungal infection or from autoimmunity, cancer, allergy, asthma,
CC   graft-versus-host disease, eczema, haemophilia, thrombosis,
CC   anti-inflammatory diseases, nervous system disorders, and infection.
CC   The present sequence encodes a protein of the invention.
XX
SQ   Sequence 3690 BP; 925 A; 947 C; 956 G; 862 T; 0 other;

                           Qy  1649 accg 1652
                                     ||||
                           Db  1485 agcg 1488

                           RESULT 15
                           AAC46946
                           ID   AAC46946 standard; DNA; 1836 BP.
                           XX
                           AC   AAC46946;
                           XX
                           DT   18-OCT-2000  (first entry)
                           XX
                           DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 51990.
                           XX
                           KW   Hybridisation assay; genetic mapping; gene expression control;
                           KW   protein identification; signal transduction pathway;
                           KW   metabolic pathway; promoter; termination sequence; ss.
                           XX
                           OS   Arabidopsis thaliana.
                           XX
                           PN   EP1033405-A2.
                           XX
                           PD   06-SEP-2000.
                           XX
                           PF   25-FEB-2000; 2000EP-0301439.
                           XX
                           PR   25-FEB-1999; 99US-0121825.
                           PR   05-MAR-1999; 99US-0123180.
                           PR   09-MAR-1999; 99US-0123548.
                           PR   23-MAR-1999; 99US-0125788.
                           PR   25-MAR-1999; 99US-0126264.
                           PR   29-MAR-1999; 99US-0126785.
                           PR   01-APR-1999; 99US-0127462.
                           PR   06-APR-1999; 99US-0128234.
                           PR   08-APR-1999; 99US-0128714.
                           PR   16-APR-1999; 99US-0129845.
                           PR   19-APR-1999; 99US-0130077.
                           PR   21-APR-1999; 99US-0130449.
                           PR   23-APR-1999; 99US-0130510.
                           PR   23-APR-1999; 99US-0130891.
                           PR   28-APR-1999; 99US-0131449.
                           PR   30-APR-1999; 99US-0132048.
                           PR   30-APR-1999; 99US-0132407.
                           PR   04-MAY-1999; 99US-0132484.
                           PR   05-MAY-1999; 99US-0132485.
                           PR   06-MAY-1999; 99US-0132486.
                           PR   06-MAY-1999; 99US-0132487.
                           PR   07-MAY-1999; 99US-0132863.
                           PR   11-MAY-1999; 99US-0134256.
                           PR   14-MAY-1999; 99US-0134218.
                           PR   14-MAY-1999; 99US-0134219.
                           PR   14-MAY-1999; 99US-0134221.
                           PR   14-MAY-1999; 99US-0134370.
                           PR   18-MAY-1999; 99US-0134768.
                           PR   19-MAY-1999; 99US-0134941.
                           PR   20-MAY-1999; 99US-0135124.
                           PR   21-MAY-1999; 99US-0135353.
                           PR   24-MAY-1999; 99US-0135629.
                           PR   25-MAY-1999; 99US-0136021.
                           PR   27-MAY-1999; 99US-0136392.
                           PR   28-MAY-1999; 99US-0136782.
                           PR   01-JUN-1999; 99US-0137222.
                           PR   03-JUN-1999; 99US-0137528.
                           PR   04-JUN-1999; 99US-0137502.
                           PR   07-JUN-1999; 99US-0137724.
                           PR   08-JUN-1999; 99US-0138094.
                           PR   10-JUN-1999; 99US-0138540.
                           PR   10-JUN-1999; 99US-0138847.
                           PR   14-JUN-1999; 99US-0139119.
                           PR   16-JUN-1999; 99US-0139452.
                           PR   16-JUN-1999; 99US-0139453.
```

```
Query Match             2.4%;     Score 45.6;  DB 22;    Length 3690;
Best Local Similarity  49.2%;     Pred. No. 0.011;
Matches  120;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps     0;

Qy  1409 actttgatccagtagccgcccgcatgcaacacctggatctgttcacctggatgcgcggcgaca 1468
          ||||||  |||||||| ||||||||||||||||| |||| ||| ||||||||||| ||||||
Db  1245 actttggccgtctggtcaacctccagcacctggatctcctccaacaacaagctggtcacct 1304

Qy  1469 tcctggtcaaggctgacaagatcaacatgcgaactccctgactgcgagttccattct 1528
          |  |||||  ||||||| || || ||||| |||||| |||||  ||| |||||
Db  1305 tgcctgtcagctttgctcagctcaagaacctgaagtggttggacctgaaggataaccccc 1364

Qy  1529 tggataaggaagttttcaagttgcagagaccattccttacgatctgaagattgccaacg 1588
          |||||  |||||| ||||||| |||||  ||||| | || || |||||||| |
Db  1365 tggatccctgccggccaaggtggcaggtggctgcttggatgagaagcagtgtaagcagt 1424

Qy  1589 gtaccaagtacgcgctgcgcaggcactcgagcacgagcagattgttccgcctcacgttttgc 1648
          |||||  |||||  |  ||  ||||||| | ||||||||  ||| |||| |
Db  1425 gtgcaaacaaggtgttacagcacatcgaaggccgtgcaggccgtgcaggatccaggaggacgcgggagaggc 1484
```

```
PR  17-JUN-1999;  99US-0139492.          PR  13-AUG-1999;  99US-0148684.
PR  18-JUN-1999;  99US-0139454.          PR  16-AUG-1999;  99US-0149368.
PR  18-JUN-1999;  99US-0139455.          PR  17-AUG-1999;  99US-0149175.
PR  18-JUN-1999;  99US-0139456.          PR  18-AUG-1999;  99US-0149426.
PR  18-JUN-1999;  99US-0139457.          PR  20-AUG-1999;  99US-0149722.
PR  18-JUN-1999;  99US-0139458.          PR  20-AUG-1999;  99US-0149723.
PR  18-JUN-1999;  99US-0139459.          PR  20-AUG-1999;  99US-0149929.
PR  18-JUN-1999;  99US-0139460.          PR  23-AUG-1999;  99US-0149902.
PR  18-JUN-1999;  99US-0139461.          PR  23-AUG-1999;  99US-0149930.
PR  18-JUN-1999;  99US-0139462.          PR  25-AUG-1999;  99US-0150566.
PR  18-JUN-1999;  99US-0139463.          PR  26-AUG-1999;  99US-0150884.
PR  18-JUN-1999;  99US-0139750.          PR  27-AUG-1999;  99US-0151065.
PR  18-JUN-1999;  99US-0139763.          PR  27-AUG-1999;  99US-0151066.
PR  21-JUN-1999;  99US-0139817.          PR  27-AUG-1999;  99US-0151080.
PR  22-JUN-1999;  99US-0139899.          PR  30-AUG-1999;  99US-0151303.
PR  23-JUN-1999;  99US-0140353.          PR  31-AUG-1999;  99US-0151438.
PR  23-JUN-1999;  99US-0140354.          PR  01-SEP-1999;  99US-0151930.
PR  24-JUN-1999;  99US-0140695.          PR  07-SEP-1999;  99US-0152363.
PR  28-JUN-1999;  99US-0140823.          PR  10-SEP-1999;  99US-0153070.
PR  29-JUN-1999;  99US-0140991.          PR  13-SEP-1999;  99US-0153758.
PR  30-JUN-1999;  99US-0141287.          PR  15-SEP-1999;  99US-0154018.
PR  01-JUL-1999;  99US-0141842.          PR  16-SEP-1999;  99US-0154039.
PR  01-JUL-1999;  99US-0142154.          PR  20-SEP-1999;  99US-0154779.
PR  02-JUL-1999;  99US-0142055.          PR  22-SEP-1999;  99US-0155139.
PR  06-JUL-1999;  99US-0142390.          PR  23-SEP-1999;  99US-0155486.
PR  08-JUL-1999;  99US-0142803.          PR  24-SEP-1999;  99US-0155659.
PR  09-JUL-1999;  99US-0142920.          PR  28-SEP-1999;  99US-0156458.
PR  12-JUL-1999;  99US-0142977.          PR  29-SEP-1999;  99US-0156596.
PR  13-JUL-1999;  99US-0143542.          PR  04-OCT-1999;  99US-0157117.
PR  14-JUL-1999;  99US-0143624.          PR  05-OCT-1999;  99US-0157753.
PR  15-JUL-1999;  99US-0144005.          PR  06-OCT-1999;  99US-0157865.
PR  16-JUL-1999;  99US-0144085.          PR  07-OCT-1999;  99US-0158029.
PR  16-JUL-1999;  99US-0144086.          PR  08-OCT-1999;  99US-0158232.
PR  19-JUL-1999;  99US-0144325.          PR  12-OCT-1999;  99US-0158369.
PR  19-JUL-1999;  99US-0144331.          PR  13-OCT-1999;  99US-0159293.
PR  19-JUL-1999;  99US-0144332.          PR  13-OCT-1999;  99US-0159294.
PR  19-JUL-1999;  99US-0144333.          PR  13-OCT-1999;  99US-0159295.
PR  19-JUL-1999;  99US-0144334.          PR  14-OCT-1999;  99US-0159329.
PR  19-JUL-1999;  99US-0144335.          PR  14-OCT-1999;  99US-0159330.
PR  20-JUL-1999;  99US-0144352.          PR  14-OCT-1999;  99US-0159331.
PR  20-JUL-1999;  99US-0144632.          PR  14-OCT-1999;  99US-0159637.
PR  20-JUL-1999;  99US-0144884.          PR  14-OCT-1999;  99US-0159638.
PR  21-JUL-1999;  99US-0144814.          PR  18-OCT-1999;  99US-0159584.
PR  22-JUL-1999;  99US-0145086.          PR  21-OCT-1999;  99US-0160741.
PR  22-JUL-1999;  99US-0145088.          PR  21-OCT-1999;  99US-0160767.
PR  22-JUL-1999;  99US-0145085.          PR  21-OCT-1999;  99US-0160768.
PR  22-JUL-1999;  99US-0145087.          PR  21-OCT-1999;  99US-0160770.
PR  22-JUL-1999;  99US-0145089.          PR  21-OCT-1999;  99US-0160814.
PR  23-JUL-1999;  99US-0145192.          PR  21-OCT-1999;  99US-0160815.
PR  23-JUL-1999;  99US-0145145.          PR  22-OCT-1999;  99US-0160980.
PR  23-JUL-1999;  99US-0145218.          PR  22-OCT-1999;  99US-0160981.
PR  23-JUL-1999;  99US-0145224.          PR  22-OCT-1999;  99US-0160989.
PR  26-JUL-1999;  99US-0145276.          PR  25-OCT-1999;  99US-0161404.
PR  27-JUL-1999;  99US-0145913.          PR  25-OCT-1999;  99US-0161405.
PR  27-JUL-1999;  99US-0145918.          PR  25-OCT-1999;  99US-0161406.
PR  27-JUL-1999;  99US-0145919.          PR  26-OCT-1999;  99US-0161359.
PR  28-JUL-1999;  99US-0145951.          PR  26-OCT-1999;  99US-0161360.
PR  02-AUG-1999;  99US-0146386.          PR  26-OCT-1999;  99US-0161361.
PR  02-AUG-1999;  99US-0146388.          PR  28-OCT-1999;  99US-0161920.
PR  03-AUG-1999;  99US-0146389.          PR  28-OCT-1999;  99US-0161992.
PR  03-AUG-1999;  99US-0147038.          PR  28-OCT-1999;  99US-0161993.
PR  04-AUG-1999;  99US-0147204.          PR  29-OCT-1999;  99US-0162142.
PR  05-AUG-1999;  99US-0147302.
PR  05-AUG-1999;  99US-0147192.
PR  06-AUG-1999;  99US-0147260.
PR  06-AUG-1999;  99US-0147303.
PR  06-AUG-1999;  99US-0147416.
PR  09-AUG-1999;  99US-0147493.
PR  09-AUG-1999;  99US-0147935.
PR  10-AUG-1999;  99US-0148171.
PR  11-AUG-1999;  99US-0148319.
PR  12-AUG-1999;  99US-0148341.
PR  13-AUG-1999;  99US-0148565.
```

```
Query Match               2.3%;      Score 44.4; DB 21; Length 1836;
Best Local Similarity    46.4%;      Pred. No. 0.018;
Matches 233; Conservative 46.4%; Mismatches 251; Indels 18; Gaps 2;

Qy    1 atgtgcggcccttcttgcatattgactgcaaatggaacgctgaagcattcgttcctgca  60
          ||||| || ||| || || |||     |  ||   | |||| || |    || ||
Db    1 atgtgtgggattctcgctgttcttggttcatcgacaactccaagctaaacgttctcgt  60

Qy   61 ctcgagcggcgggcccttgccatgccatgcatgccgccaccgtggtcctgacgatgccggccacttggcat 120
          ||    |||   ||| ||||| ||||     ||||    ||||||    |||||||
```

```
Db      61 atcatcggaactctctcgcagattgaggcacagaggtcctgattggagtgactccattgt 120

Qy     121 gacgccgatgcagcgtttggattcaaccgcctctccatcattgatattgcacactcccac 180
              |--|--|||    -||-|    |||||||||  |-|              ||
Db     121 tatgaagattgttatcttgcccatgagcgttagccatcattgacccctacttcaggagac 180

Qy     181 caaccactgcgttggggacctgcgatgaaccgaccgctacgcaatgactttcaacggt 240
              ||||||  |-|           |-|||-|  |||-|  |||||| | ||-|||
Db     181 caacctctctat---------aacggagacaagaccgtcgctgtcactgtaaatgga 228

Qy     241 gagatctacaactacgttgagctgcgtaaagagctctcggatttgggatatacctttaat 300
              |||||  ||||||| ||-|||||  |||||||||  |-||     ||  |-|  ||
Db     229 gagatatacaacaccacaagatttttgcgtgaaaagct-----taagtctcatcagttccgt 282

Qy     301 acttctggcgatgcgagccaattgttgtcggttccaccactgggcgagtccgtggtc 360
              ||--|            |||||          |-|||   |-|||    -|--||
Db     283 actggtagtgactgtgaagtgattgcacatcttacggagaacatggagaggaatttatc 342

Qy     361 gagcatctccgcggaatgttcggcattgccattggatacaaaggaaaagtcgctttc 420
              ||--|   ||||||||||  |||-|||  |-||    |||||  ||   ||-||
Db     343 gacatgttggatggaatgttcgcgtttgtccttcttgatactgcgcgacaaaagttttatt 402

Qy     421 cttgcgcgtgatcagttcggcatcaagccactgttctacgcaccaccgagcatggcacc 480
              |||||||    -|-|-|-||  |||||  |||-|-|    |-|    ||||
Db     403 gctgcaaggggacgctattggtatcactccacttacattggatgggtcttgatggttct 462

Qy     481 gtgttctctcagagaagaaga 502
              ||-|||    ||-|||||||
Db     463 gtctggtttgctctcggagatga 484
```

Search completed: September 20, 2002, 07:36:04
Job time: 6051 sec

Copyright GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 20, 2002, 05:54:08 ; Search time 58.98 Seconds
                                               (without alignments)
                                               7996.208 Million cell updates/sec

Title: US-09-786-474-1
Perfect score: 1920
Sequence: 1 atgtgcggcctctcttggcat.........gctcctacccggtcgagctt 1920

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_NA:*
                 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
                 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
                 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
                 6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match Score	Length	DB	ID	Description
1	116.6	6.1	30001	1	US-08-125-468-1	Sequence 1, Appli
2	116.6	6.1	30001	1	US-08-474-933-1	Sequence 1, Appli
3	37.6	2.0	1349	4	US-09-191-608-15	Sequence 15, Appl
4	37.6	2.0	1421	4	US-09-191-608-14	Sequence 14, Appl
5	37.6	2.0	1436	4	US-09-191-608-13	Sequence 13, Appl
6	37.6	2.0	1499	4	US-09-191-608-16	Sequence 16, Appl
7	36.4	1.9	1542	4	US-09-385-028-16	Sequence 16, Appl
8	36.4	1.9	11604	4	US-09-385-028-13	Sequence 13, Appl
9	36.4	1.9	15079	4	US-09-385-028-1	Sequence 1, Appli
10	36	1.9	531	4	US-09-191-608-8	Sequence 8, Appli
11	34.8	1.8	2267	4	US-08-679-645-25	Sequence 25, Appl
12	34.4	1.8	323	2	US-08-997-080-92	Sequence 92, Appl
13	34.4	1.8	323	3	US-08-997-362-92	Sequence 92, Appl
14	34.4	1.8	323	3	US-08-873-970-92	Sequence 92, Appl
15	34.4	1.8	323	4	US-09-095-855-92	Sequence 92, Appl
16	34.4	1.8	323	4	US-09-324-542-92	Sequence 92, Appl
17	34.4	1.8	1111	2	US-08-997-080-169	Sequence 169, App
18	34.4	1.8	1111	3	US-08-997-362-169	Sequence 169, App
19	34.4	1.8	1111	4	US-09-095-855-169	Sequence 169, App
20	34.4	1.8	1111	4	US-09-324-542-169	Sequence 169, App
21	34.4	1.8	1341	2	US-08-997-080-93	Sequence 93, Appl
22	34.4	1.8	1341	3	US-08-997-362-93	Sequence 93, Appl
23	34.4	1.8	1341	3	US-08-873-970-93	Sequence 93, Appl
24	34.4	1.8	1341	4	US-09-095-855-93	Sequence 93, Appl
25	34.4	1.8	1341	4	US-09-324-542-93	Sequence 93, Appl
26	34.4	1.8	2184	4	US-09-307-265A-2	Sequence 2, Appli
27	34.2	1.8	40138	3	US-09-090-793-12	Sequence 12, Appl
c 28	34.2	1.8	4411529	4	US-09-103-840A-1	Sequence 1, Appli
29	34	1.8	1050	4	US-09-655-270A-16	Sequence 16, Appl
30	34	1.8	1050	4	US-09-651-941-20	Sequence 20, Appl
31	34	1.8	1886	6	5256558-1	Patent No. 5256558
32	34	1.8	12508	4	US-09-655-270A-1	Sequence 1, Appli
33	34	1.8	12523	4	US-09-651-941-1	Sequence 1, Appli
34	33.8	1.8	11459	4	US-09-454-721A-3	Sequence 3, Appli
35	33.6	1.8	3624	4	US-08-434-730-13	Sequence 13, Appl
c 36	33.6	1.8	4403765	4	US-09-103-840A-2	Sequence 2, Appli
37	33.4	1.7	2160	4	US-09-588-256-1	Sequence 1, Appli
38	33	1.7	1026	1	US-07-975-526-6	Sequence 6, App
39	33	1.7	1026	4	US-07-974-409C-428	Sequence 428, App
40	32.8	1.7	4689	4	US-09-105-537-34	Sequence 34, Appl
41	32.8	1.7	36778	3	US-09-105-537-5	Sequence 5, Appli
42	32.8	1.7	38506	3	US-09-320-878-19	Sequence 19, Appl
43	32.4	1.7	1220	3	US-09-148-801-1	Sequence 1, Appli
c 44	32.2	1.7	710	3	US-09-098-287A-5	Sequence 5, Appli
45	32.2	1.7	3301	2	US-08-447-430A-42	Sequence 42, Appl

ALIGNMENTS

RESULT 1
US-08-125-468-1/c
; Sequence 1, Application US/08125468
; Patent No. 5589385
; GENERAL INFORMATION:
; APPLICANT: Ryan, Michael J.
; APPLICANT: Lotvin, Jason A.
; APPLICANT: Strathy, Nancy
; APPLICANT: Fantini, Susan E.
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for
; TITLE OF INVENTION: chlortetracycline and tetracyline Formation and cosmids
; TITLE OF INVENTION: useful therein
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: American Cyanamid Company
;   STREET: One Cyanamid Plaza
;   CITY: Wayne
;   STATE: New Jersey
;   COUNTRY: USA
;   ZIP: 07470
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/125,468
;   FILING DATE: 22-SEP-1993
;   CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
;   NAME: Tsevdos, Estelle J
;   REGISTRATION NUMBER: 31,145
;   REFERENCE/DOCKET NUMBER: 31,255-02
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (201)831-3241
;   TELEFAX: (201)831-3305
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 30001 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-08-125-468-1

Query Match   6.1%; Score 116.6; DB 1; Length 30001;
Best Local Similarity   56.1%; Pred. No. 1.4e-24;
Matches 243; Conservative   0; Mismatches 184; Indels   6; Gaps   1;

```
  SEQUENCE CHARACTERISTICS:
    LENGTH: 30001 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
    MOLECULE TYPE: DNA (genomic)
US-08-474-933-1

Query Match
Best Local Similarity   56.1%;              Score 116.6;  DB 2;  Length 30001;
Matches  243;  Conservative  0;             Pred. No. 1.4e-24;
               Mismatches 184;  Indels   6;  Gaps   1;

Qy    81  catgcgccaccgtggtcctgacgatgccggcatgacgccgatgcagcgttgg        140
          |||| | ||| |||| |||  || || || |||||| |||||||||||||
Db 22834  CATGGCCTGCCGGGGCCGGACGGCCCTCTGGTTCGGCCGCGGCGGCCTCGG       22775

Qy   141  attcaaccgcctctccatcattgatattgcacactcccaccaccactgcgttgggggacc  200
          | |||| ||| | ||| ||||| ||  |  ||| ||||| ||  ||| || ||  ||
Db 22774  TCACCGGCGCGTGTCGGTGATCGACCCCGAGCACGGCCCGAGTGACCGCGGAAC-      22714

Qy   201  tcggatgaacccgacccgctacgcaatgactttcaacggtgagatctacaactacgttga  260
          ---||| || ||| |||||||| || ||| |||||| || || |||| |||| |  |
Db 22715  -----ACGAGGGCGCCGGCCGGGCGGATCACCTTCAGCGCGCGAGAATCTACAACTTCCGTGA 22661

Qy   261  gctgcgtaaagagctccggatttgggatataccttataatacttctggcgatggcgagcc  320
          |||||| | |  || ||| | |||||| || ||   ||| ||| ||||| || |||
Db 22660  GCTGCGCGCCGAACTCACCTCGCACCGGCCACCGGTTCAGGACGTCCTGCGATACCGAGGT 22601

Qy   321  aattgtgtcggttccaccactggggcgagtccgtggtcgagcatctccgcggaatgtt    380
          || | || || ||| ||| ||||| || ||||| ||  ||  |  | || || ||
Db 22600  CGTCCTGCGCGGCTACCTCCAGTGGGCGCCGCCGCTGGGTCGTCGACGGCTCAACGGGATGT 22541

Qy   381  cggcattgcattgggatacaaaggaaaagtcgctttccttgcgcgtgatcagttcgg     440
          ||| || ||| |||||| || || ||||  || ||||| || |  |||| || ||
Db 22540  CGCCTTCGCCGTCTGGGACGAGCACAGCGAGGAACTCCTGCTGGTCCCGCACCGGATGGG  22481

Qy   441  catcaagccactgttctacgcaaccaccgagcatgcaccgtgttctcctcagagaagaa   500
          || ||| | || ||||| || ||||||| || || ||| ||| ||||| ||||| ||
Db 22480  CGTCAAGCGCGCTCTACTACTGGCCGACCGCCGCCGGCGTGCCGCTTCGGCTCGGAGCCCAA 22421

Qy   501  gaccatcttggag  513
          || ||||| |
Db 22420  GGCCGTCCTGGCG  22408
```

RESULT 2
US-08-474-933-1/c
; Sequence 1, Application US/08474933
; Patent No. 5866410
; GENERAL INFORMATION:
;   APPLICANT: Ryan, Michael J.
;   APPLICANT: Lotvin, Jason A.
;   APPLICANT: Strathy, Nancy
;   APPLICANT: Fantini, Susan E.
;   TITLE OF INVENTION: Cloning of the biosynthetic pathway for
;   TITLE OF INVENTION: chlortetracycline and tetracycline Formation and cosmids
;   TITLE OF INVENTION: useful therein
;   NUMBER OF SEQUENCES: 1
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: American Cyanamid Company
;   STREET: One Cyanamid Plaza
;   CITY: Wayne
;   STATE: New Jersey
;   COUNTRY: USA
;   ZIP: 07470
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/474,933
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/125,468
;   FILING DATE: 22-SEP-1993
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Tsevdos, Estelle J
;   REGISTRATION NUMBER: 31,145
;   REFERENCE/DOCKET NUMBER: 31,255-02
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (201)831-3241
;   TELEFAX: (201)831-3305
; INFORMATION FOR SEQ ID NO: 1:

RESULT 3
US-09-191-608-15/c
; Sequence 15, Application US/0919608
; Patent No. 6242216
; GENERAL INFORMATION:
;   APPLICANT: Lynch, Kevin J.
;   APPLICANT: Burgard, Edward C.
;   APPLICANT: Metzger, Randy E.
;   APPLICANT: Niforatos, Wende
;   APPLICANT: Touma, Edward B.
;   APPLICANT: Van Biesen, T.
;   TITLE OF INVENTION: Nucleic Acids Encoding a Functional
;   TITLE OF INVENTION: Human Purinoreceptor P2X2 and P2X4 And Methods Of Production
;   TITLE OF INVENTION: And Use Thereof
;   FILE REFERENCE: 6394.US.P1
;   CURRENT APPLICATION NUMBER: US/09/191,608
;   CURRENT FILING DATE: 1998-11-13
;   NUMBER OF SEQ ID NOS: 26
;   SOFTWARE: FasteSEQ for Windows Version 3.0
; SEQ ID NO 15
;   LENGTH: 1349
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-191-608-15

Query Match   2.0%; Score 37.6; DB 4; Length 1349;

Best Local Similarity 47.4%; Pred. No. 0.19;
Matches 145; Conservative 0; Mismatches 159; Indels 2; Gaps 1;

```
Qy    286 ggatatacctttaatacttctggcgatggcgagccaattgtgtcggttccaccactgg 345
              |||||  ||  | || ||||  ||  || |  ||||| | ||| |||||||||||
Db    872 GGTGGTGCCATTGATCTTGTAGTATTTGGCAAACCTGAAGTTGTAGCCTGACGAGGCAGG 813

Qy    346 ggcgagtccgtggtcgagcatctccgcggaatgttcgcattgggatacaaag 405
              ||||| || | || ||||   ||| | | || || |  ||| ||||| ||
Db    812 CACGTGCTTGGGGTCAAGCCTCCGAAGGAGTACTTGGGTGCACTCCGATGCAGGCAG 753

Qy    406 gaaaagtcgctttccttgcgcgtgatcagttcggcatcaagccactgttctacgcaacc 465
              ||  |||  |||||| ||| | || |  || ||| ||| | ||  || ||  ||
Db    752 GTCCAGGTCACAGTCCCAGTTGATAATGACCCCG--ATGACACCACCCTTGTGTGCGAGC 695

Qy    466 accgagcatggcaccgtgttcctccagagaagaccatcttggagatggccgaggag 525
              |||||||| |||| | |  |||| | || | |||| || ||||  || ||||||
Db    694 TCTGTGAAGCTCCCCCAGCCTTCTCCACGATAAAGCCCAGCTTGAAGATGGGGCAGTAG 635

Qy    526 atgaatctagatctgggcctgataagcgccattgagcactacgtggacctgcagtac 585
              ||||| |  ||||||| || |||| ||||| ||  || ||||||| |||| ||||
Db    634 AGTTCGGAGGCCTTCGTGGAACGTGCAGCGCCTTCAGGTACCCGTCTGTGCGGTCGGCGATG 575

Qy    586 gtgccc 591
              |||||
Db    574 TTGCCC 569
```

RESULT 4
US-09-191-608-14/c
; Sequence 14, Application US/09191608
; Patent No. 6242216
; GENERAL INFORMATION:
; APPLICANT: Lynch, Kevin J.
; APPLICANT: Burgard, Edward C.
; APPLICANT: Metzger, Randy E.
; APPLICANT: Niforatos, Wende
; APPLICANT: Touma, Edward B.
; APPLICANT: Van Biesen, T.
; TITLE OF INVENTION: Nucleic Acids Encoding a Functional
; TITLE OF INVENTION: Human Purinoreceptor P2X2 and P2X4 And Methods Of Production
; TITLE OF INVENTION: And Use Thereof
; FILE REFERENCE: 6394.US.P1
; CURRENT APPLICATION NUMBER: US/09/191,608
; CURRENT FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 1421
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-191-608-14

Query Match 2.0%; Score 37.6; DB 4; Length 1421;
Best Local Similarity 47.4%; Pred. No. 0.2;
Matches 145; Conservative 0; Mismatches 159; Indels 2; Gaps 1;

```
Qy    286 ggatatacctttaatacttctggcgatggcgagccaattgtgtcggttccaccactgg 345
              |||||  ||  | || ||||  ||  || |  ||||| | ||| |||||||||||
Db    944 GGTGGTGCCATTGATCTTGTAGTATTTGGCAAACCTGAAGTTGTAGCCTGACGAGGCAGG 885

Qy    346 ggcgagtccgtggtcgagcatctccgcggaatgttcgcattgggatacaaag 405
              ||||| || | || ||||   ||| | | || || |  ||| ||||| ||
Db    884 CACGTGCTTGGGGTCAAGCCTCCGAAGGAGTACTTGGGTGCACTCCGATGCAGGCAG 825

Qy    406 gaaaagtcgctttccttgcgcgtgatcagttcggcatcaagccactgttctacgcaacc 465
              ||  |||  |||||| ||| | || |  || ||| ||| | ||  || ||  ||
Db    824 GTCCAGGTCACAGTCCCAGTTGATAATGACCCCG--ATGACACCACCCTTGTGTGCGAGC 767

Qy    466 accgagcatggcaccgtgttcctccagagaagaccatcttggagatggccgaggag 525
              |||||||| |||| | |  |||| | || | |||| || ||||  || ||||||
Db    766 TCTGTGAAGCTCCCCCAGCCTTCTCCACGATAAAGCCCAGCTTGAAGATGGGGCAGTAG 707
```

```
Qy    526 atgaatctagatctgggcctgataagcgccattgagcactacgtggacctgcagtac 585
              ||||| |  ||||||| || |||| ||||| ||  || ||||||| |||| ||||
Db    706 AGTTCGGAGGCCTTCGTGGAACGTGCAGCGCCTTCAGGTACCCGTCTGTGCGGTCGGCGATG 647

Qy    586 gtgccc 591
              |||||
Db    646 TTGCCC 641
```

RESULT 5
US-09-191-608-13/c
; Sequence 13, Application US/09191608
; Patent No. 6242216
; GENERAL INFORMATION:
; APPLICANT: Lynch, Kevin J.
; APPLICANT: Burgard, Edward C.
; APPLICANT: Metzger, Randy E.
; APPLICANT: Niforatos, Wende
; APPLICANT: Touma, Edward B.
; APPLICANT: Van Biesen, T.
; TITLE OF INVENTION: Nucleic Acids Encoding a Functional
; TITLE OF INVENTION: Human Purinoreceptor P2X2 and P2X4 And Methods Of Production
; TITLE OF INVENTION: And Use Thereof
; FILE REFERENCE: 6394.US.P1
; CURRENT APPLICATION NUMBER: US/09/191,608
; CURRENT FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 1436
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-191-608-13

Query Match 2.0%; Score 37.6; DB 4; Length 1436;
Best Local Similarity 47.4%; Pred. No. 0.2;
Matches 145; Conservative 0; Mismatches 159; Indels 2; Gaps 1;

```
Qy    286 ggatatacctttaatacttctggcgatggcgagccaattgtgtcggttccaccactgg 345
              |||||  ||  | || ||||  ||  || |  ||||| | ||| |||||||||||
Db    944 GGTGGTGCCATTGATCTTGTAGTATTTGGCAAACCTGAAGTTGTAGCCTGACGAGGCAGG 885

Qy    346 ggcgagtccgtggtcgagcatctccgcggaatgttcgcattgggatacaaag 405
              ||||| || | || ||||   ||| | | || || |  ||| ||||| ||
Db    884 CACGTGCTTGGGGTCAAGCCTCCGAAGGAGTACTTGGGTGCACTCCGATGCAGGCAG 825

Qy    406 gaaaagtcgctttccttgcgcgtgatcagttcggcatcaagccactgttctacgcaacc 465
              ||  |||  |||||| ||| | || |  || ||| ||| | ||  || ||  ||
Db    824 GTCCAGGTCACAGTCCCAGTTGATAATGACCCCG--ATGACACCACCCTTGTGTGCGAGC 767

Qy    466 accgagcatggcaccgtgttcctccagagaagaccatcttggagatggccgaggag 525
              |||||||| |||| | |  |||| | || | |||| || ||||  || ||||||
Db    766 TCTGTGAAGCTCCCCCAGCCTTCTCCACGATAAAGCCCAGCTTGAAGATGGGGCAGTAG 707

Qy    526 atgaatctagatctgggcctgataagcgccattgagcactacgtggacctgcagtac 585
              ||||| |  ||||||| || |||| ||||| ||  || ||||||| |||| ||||
Db    706 AGTTCGGAGGCCTTCGTGGAACGTGCAGCGCCTTCAGGTACCCGTCTGTGCGGTCGGCGATG 647

Qy    586 gtgccc 591
              |||||
Db    646 TTGCCC 641
```

RESULT 6
US-09-191-608-16/c
; Sequence 16, Application US/09191608
; Patent No. 6242216
; GENERAL INFORMATION:
; APPLICANT: Lynch, Kevin J.
; APPLICANT: Burgard, Edward C.
; APPLICANT: Metzger, Randy E.

```
;     APPLICANT: Niforatos, Wende
;     APPLICANT: Touma, Edward B.
;     APPLICANT: Van Biesen, T.
;     TITLE OF INVENTION: Nucleic Acids Encoding a Functional
;     TITLE OF INVENTION: Human Purinoreceptor P2X2 and P2X4 And Methods Of Production
;     TITLE OF INVENTION: And Use Thereof
;     FILE REFERENCE: 6394.US.P1
;     CURRENT APPLICATION NUMBER: US/09/191,608
;     CURRENT FILING DATE: 1998-11-13
;     NUMBER OF SEQ ID NOS: 26
;     SOFTWARE: FastSEQ for Windows Version 3.0
;     SEQ ID NO 16
;     LENGTH: 1499
;     TYPE: DNA
;     ORGANISM: Homo sapiens
US-09-191-608-16

Query Match          2.0%;     Score 37.6;   DB 4;   Length 1499;
Best Local Similarity  47.4%;   Pred. No. 0.2;   Indels   2;   Gaps   1;
Matches   145; .Conservative    0;   Mismatches  159;   Indels   0;

         286 ggatatacctttaatacttctggcgatggcgagccagccaattgttgtcggttccaccactgg 345
             ||   |||   |       |   ||| ||  || | ||    |   ||   ||  ||| ||
Db       944 GGTGGTGCCATTGATCTTGTAGTATTTGGCAAACCTGAAGTTGTAGCCTGACGAGGCAGG 885

Qy       346 ggcgagtccggtcgagcatctccgcggaatgttcgcattgccatttgggatacaaag 405
             ||   || ||  || || |      |||  ||     ||| |  ||  ||||
Db       884 CACGTGCTTGGGGTCAAGCCTCCGGAAGGAGTACTTGGGGTGCACTCCGATGCAGGCAG 825

Qy       406 gaaaagtcgctttcttcttgcgcgtgatcagttcggcatcaagccactgttctacgcaacc 465
             ||    |  | |||  |  ||  || ||||  |   ||| |||| |||  ||
Db       824 GTCCAGTCACAGTCCCAGTTGATAATGACCCCG--ATGACACCACCCTTGTGTGCGAGC 767

Qy       466 accgagcatggcaccgtgtctcctccagagaagaagaccatcttggagatggccgaggag 525
             ||| || ||   || ||  |  |   || || || |    ||  |||| | |||   |
Db       766 TCTGTGAAGCTCTCCCCAGCCTTCTCCACGATAAAGCCCAGCTTGAAGATGGGGCAGTAG 707

Qy       526 atgaatctagatctggccttgataagcgcaccattgagcactacgtggacctgcagtac 585
             ||    |  || |  ||  || ||||| ||   ||| ||| |  |||| |||  ||
Db       706 AGTTCGGAGCCAGCCTCGTGGAACGTGCAGCCGCTTCAGGTACCCGTCTGTGCCGTCGGCGATG 647

Qy       586 gtgccc 591
             |||||
Db       646 TTGCCC 641


RESULT  7
US-09-385-028-16
;     Sequence 16, Application US/09385028
      Patent No. 6232106
;     GENERAL INFORMATION:
;     APPLICANT: Susan E. Jensen
;     APPLICANT: Kwamena A Aidoo
;     APPLICANT: Ashish S. Paradkar
;     TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
      Patent No. 6232106
;     TITLE OF INVENTION: Acid Biosynthesis
;     NUMBER OF SEQUENCES: 25
;     CORRESPONDENCE ADDRESS:
;     ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN,PPLC
;     STREET: The Jenifer Building, 400 Seventh Street, N.W.
;     CITY: Washington
;     STATE: D.C.
;     COUNTRY: U.S.A.
;     ZIP: 20004
;     COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
;     CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/385,028
```

```
;     FILING DATE:
;     CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/790,462
;     FILING DATE:   29-JAN-1997
;     ATTORNEY/AGENT INFORMATION:
;     NAME: D. Douglas Price
;     REGISTRATION NUMBER: 24,514
;     REFERENCE/DOCKET NUMBER: 1418/P57452US2
;     TELECOMMUNICATION INFORMATION:
;     TELEPHONE:   (202) 638-6666
;     TELEFAX:   (202) 39305350
;     TELEX:   RCA 248593 IDEA UR
;     INFORMATION FOR SEQ ID NO: 16:
;     SEQUENCE CHARACTERISTICS:
;     LENGTH: 1542 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;     MOLECULE TYPE: DNA (genomic)
US-09-385-028-16

Query Match          1.9%;     Score 36.4;   DB 4;   Length 1542;
Best Local Similarity  50.6%;   Pred. No. 0.47;   Indels   0;   Gaps   0;
Matches   88;   Conservative    0;   Mismatches  86;   Indels   0;   Gaps   0;

Qy       726 ggagcaggacctcttcgatcgcattgccagtgccagtgttggaggatagtcgaaagcatat 785
             ||  |   || || ||  ||  ||  |    ||   | | | || ||   |  |  |
Db       654 GGAGGGCGAGCCGTCGGCGGCCGCGGGCCGCGTGGCGGGCCGTCGCCCAGCGGGT 713

Qy       786 gcgtgacgacgtgaccgtaggcgtgttcctttccggccgcgcattgactaaccgcaattgc 845
             ||   | ||    |   ||||    || || |  |||  | |  ||| ||  |    ||
Db       714 CACCCCCGGCGACACCCCGTTGGTGCTCTCCGGCGGAATCGACTCCTCCGGGTCGC 773

Qy       846 gccgcttgcaaagcgccacaaccctgacctgctccacct tcaccaccggtttcga 899
             ||  || |   |||| |  || ||     ||   ||  ||  |||   |||||
Db       774 GGCCTGTGCGCACCGGGCGGCGCGGGACTGGCACGGTGTCCATGGGCACCGA 827


RESULT  8
US-09-385-028-13
;     Sequence 13, Application US/09385028
      Patent No. 6232106
;     GENERAL INFORMATION:
;     APPLICANT: Susan E. Jensen
;     APPLICANT: Kwamena A Aidoo
;     APPLICANT: Ashish S. Paradkar
;     TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
      Patent No. 6232106
;     TITLE OF INVENTION: Acid Biosynthesis
;     NUMBER OF SEQUENCES: 25
;     CORRESPONDENCE ADDRESS:
;     ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PPLC
;     STREET: The Jenifer Building, 400 Seventh Street, N.W.
;     CITY: Washington
;     STATE: D.C.
;     COUNTRY: U.S.A.
;     ZIP: 20004
;     COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
;     CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/385,028
;     FILING DATE:
;     CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/790,462
;     FILING DATE:   29-JAN-1997
;     ATTORNEY/AGENT INFORMATION:
;     NAME: D. Douglas Price
```

; ; SEQUENCE CHARACTERISTICS:
; ;   LENGTH: 15079 base pairs
; ;   TYPE: nucleic acid
; ;   STRANDEDNESS: single
; ;   TOPOLOGY: linear
; ;   MOLECULE TYPE: DNA (genomic)
; ; HYPOTHETICAL: NO
; ; ORIGINAL SOURCE:
; ;   ORGANISM: Streptomyces clavuligerus
US-09-385-028-1

Query Match   1.9%;   Score 36.4; DB 4; Length 15079;
Best Local Similarity 50.6%;   Pred. No. 1.7;
Matches   88; Conservative   0; Mismatches   86; Indels   0; Gaps   0;

Qy   726 ggagcaggacctcttcgatcgcattgccaggtgttggaggatagcgtcgaaaagcatat 785
          |||| || ||| || | || |||||||||||||| |||| ||||| | |||||||||
Db   4593 GGAGGGCGGAGGCCGTCCGGCGGCCGGGCCGCGTGCGGGCGGCATTGACTCAACCGCAAGGCCGTGCCCCCAGCGGGT 4652

Qy   786 gcgtcgcagtgaccgtaggctcgttccttccggcggcattgactcaaccgcaattgc 845
          ||||| || ||| ||| |||| || ||||||||||||||| ||||| ||| ||
Db   4653 CACCCCCGGCGACACCCCGTTGTGTGCTCTCCGGCGGAATCGACTCCTCCGGGGTCGC 4712

Qy   846 gccgcttgcaaagcgcacaacctgacctgtcacccttcaccaccgtttcga 899
          ||||| ||||| ||| || | | || || ||||||| || ||||||||
Db   4713 GGCCTGTGCGCACCGCACCGGCGGCGGGGAACTGGACACGGTGTCCATGGGCACCGA 4766

RESULT 10
US-09-191-608-8/c
; Sequence 8, Application US/09191608
; Patent No. 6242216
; GENERAL INFORMATION:
; ;   APPLICANT: Lynch, Kevin J.
; ;   APPLICANT: Burgard, Edward C.
; ;   APPLICANT: Metzger, Randy E.
; ;   APPLICANT: Niforatos, Wende
; ;   APPLICANT: Touma, Edward B.
; ;   APPLICANT: Van Biesen, T.
; ;   TITLE OF INVENTION: Nucleic Acids Encoding a Functional
; ;   TITLE OF INVENTION: Human Purinoreceptor P2X2 and P2X4 And Methods of Production
; ;   TITLE OF INVENTION: And Use Thereof
; ;   FILE REFERENCE: 6394.US.P1
; ;   CURRENT APPLICATION NUMBER: US/09/191,608
; ;   CURRENT FILING DATE: 1998-11-13
; ;   NUMBER OF SEQ ID NOS: 26
; ;   SOFTWARE: FastSEQ for Windows Version 3.0
; ; SEQ ID NO 8
; ;   LENGTH: 531
; ;   TYPE: DNA
; ;   ORGANISM: Homo sapiens
US-09-191-608-8

Query Match   1.9%;   Score 36; DB 4; Length 531;
Best Local Similarity 47.1%;   Pred. No. 0.35;
Matches 144; Conservative   0; Mismatches 160; Indels   2; Gaps   1;

Qy   286 ggatataccttaatacttctgcgatggcggcaagtgttgtcggtttccaccactgg 345
          || || | | | ||| ||| | | ||| | || | | || |||| | |||
Db   474 GGTGGTGCCATTGATCTTGTAGTATTTGGCAAACCTGAAGTTGTAGGCTGACGAGGGCAGG 415

Qy   346 ggcgagtccgttgtcgagcatctccgcggaatgttccggcattgccatttgggatacaaag 405
          |||| | | | || | || | |||||| ||| || ||| | |
Db   414 CACGTGCTTGGGTCAAGCCTCCGGAAGGAGTACTTGGGGTTGCACTCGATCGCAGGCAG 355

Qy   406 gaaaagtcgctttcctgcggtgatcagttcggcatcaagccactgttctacgcaacc 465
          ||| ||| || | | | | || ||| || || | | | | | || ||
Db   354 GTCCAGTCACAGTCCCAGTTGATAATGACCCG--ATGACACCACCCCTTGTGTGCCAGC 297

Qy   466 accgagcatgcaccgtgttcctcagagaagaagaacatcttgggagatggccgaggag 525
          ||| | || || || || | | | | | || | ||| ||||| |||| | ||

; ; SEQUENCE CHARACTERISTICS:
; ;   LENGTH: 11604 base pairs
; ;   TYPE: nucleic acid
; ;   STRANDEDNESS: single
; ;   TOPOLOGY: linear
; ;   MOLECULE TYPE: DNA (genomic)
US-09-385-028-13

Query Match   1.9%;   Score 36.4; DB 4; Length 11604;
Best Local Similarity 50.6%;   Pred. No. 1.4;
Matches   88; Conservative   0; Mismatches   86; Indels   0; Gaps   0;

Qy   726 ggagcaggacctcttcgatcgcattgccaggtgttggaggatagcgtcgaaaagcatat 785
          |||| || ||| || | || |||||||||||||| |||| ||||| | |||||||||
Db   2561 GGAGGGCGGAGGCCGTCCGGCGGCCGGGCCGCGTGCGGGCGGCATTGACTCAACCGCAAGGCCGTGCCCCCAGCGGGT 2620

Qy   786 gcgtcgcagtgaccgtaggctcgttccttccggcggcattgactcaaccgcaattgc 845
          ||||| || ||| ||| |||| || ||||||||||||||| ||||| ||| ||
Db   2621 CACCCCCGGCGACACCCCGTTGTGTGCTCTCCGGCGGAATCGACTCCTCCGGGGTCGC 2680

Qy   846 gccgcttgcaaagcgcacaacctgacctgtcacccttcaccaccgtttcga 899
          ||||| ||||| ||| || | | || || ||||||| || ||||||||
Db   2681 GGCCTGTGCGCACCGCACCGGCGGCGGGGAACTGGACACGGTGTCCATGGGCACCGA 2734

RESULT 9
US-09-385-028-1
; Sequence 1, Application US/09385028
; Patent No. 6232106
; GENERAL INFORMATION:
; ;   APPLICANT: Susan E. Jensen
; ;   APPLICANT: Kwamena A Aidoo
; ;   APPLICANT: Ashish S. Paradkar
; ;   TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
; Patent No. 6232106
; ;   TITLE OF INVENTION: Acid Biosynthesis
; ;   NUMBER OF SEQUENCES: 25
; ;   CORRESPONDENCE ADDRESS:
; ;   ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PPLC
; ;   STREET: The Jenifer Building, 400 Seventh Street, N.W.
; ;   CITY: Washington
; ;   STATE: D.C.
; ;   COUNTRY: U.S.A.
; ;   ZIP: 20004
; ;   COMPUTER READABLE FORM:
; ;   MEDIUM TYPE: Floppy disk
; ;   COMPUTER: IBM PC compatible
; ;   OPERATING SYSTEM: PC-DOS/MS-DOS
; ;   SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; ;   CURRENT APPLICATION DATA:
; ;   APPLICATION NUMBER: US/09/385,028
; ;   FILING DATE:
; ;   CLASSIFICATION:
; ;   PRIOR APPLICATION DATA:
; ;   APPLICATION NUMBER: US 08/790,462
; ;   FILING DATE: 29-JAN-1997
; ;   ATTORNEY/AGENT INFORMATION:
; ;   NAME: D. Douglas Price
; ;   REGISTRATION NUMBER: 24,514
; ;   REFERENCE/DOCKET NUMBER: 1418/P57452US2
; ;   TELECOMMUNICATION INFORMATION:
; ;   TELEPHONE: (202) 638-6666
; ;   TELEFAX: (202) 39305350
; ;   TELEX: RCA 248593 IDEA UR
; INFORMATION FOR SEQ ID NO: 1:

; ;   REGISTRATION NUMBER: 24,514
; ;   REFERENCE/DOCKET NUMBER: 1418/P57452US2
; ;   TELECOMMUNICATION INFORMATION:
; ;   TELEPHONE: (202) 638-6666
; ;   TELEFAX: (202) 39305350
; ;   TELEX: RCA 248593 IDEA UR
; INFORMATION FOR SEQ ID NO: 13:

```
Db    296  TCTGTGAAGCTCTCCCAGCTTCTCCACGATAAAGCCCAGCTTGAAGATGGGCAGTAA  237
Qy    526  atgaatctagatctggccttgataagcgcaccattgagcgcactactgggacctgcgagtac  585
               |-|    ||      |||||| ||  |||| ||
Db    236  AGGTCCGGAGGCCTCGTGGAACGTGCAGCGCGTTCAGTGTACCCGTCTGTGCCGTCGGCGATG  177

Qy    586  gtgccc  591
           |||||
Db    176  TTGCCCC  171
```

RESULT 11
US-08-679-645-25
; Sequence 25, Application US/08679645
; Patent No. 6350934
; GENERAL INFORMATION:
; APPLICANT: Zwick, Michael G.
; APPLICANT: Edington, Brent E.
; APPLICANT: McSwiggen, James A.
; APPLICANT: Merlo, Patricia Ann Owens
; APPLICANT: Guo, Lining
; APPLICANT: Skokut, Thomas A.
; APPLICANT: Young, Scott A.
; APPLICANT: Folkerts, Otto
; APPLICANT: Merlo, Donald J.
; TITLE OF INVENTION: COMPOSITION AND METHODS FOR
; TITLE OF INVENTION: MODULATION OF GENE EXPRESSION
; TITLE OF INVENTION: IN PLANTS
; NUMBER OF SEQUENCES: 1263
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Lyon & Lyon
;   STREET: 633 West Fifth Street
;   STREET: Suite 4700
;   CITY: Los Angeles
;   STATE: California
;   COUNTRY: U.S.A.
;   ZIP: 90071-2066
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
;   MEDIUM TYPE: storage
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: IBM P.C. DOS 5.0
;   SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/679,645
;   FILING DATE: July 12, 1996
;   CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 60/001,135
;   FILING DATE: July 13, 1995
;   APPLICATION NUMBER: 08/300,726
;   FILING DATE: September 2, 1994
; ATTORNEY/AGENT INFORMATION:
;   NAME: Warburg, Richard J.
;   REGISTRATION NUMBER: 32,327
;   REFERENCE/DOCKET NUMBER: 219/247
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (213) 489-1600
;   TELEFAX: (213) 955-0440
;   TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 2267 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
US-08-679-645-25

Query Match   1.8%;   Score 34.8;   DB 4;   Length 2267;
Best Local Similarity 45.6%;   Pred. No. 1.8;
Matches 123;   Conservative 0;   Mismatches 147;   Indels 0;   Gaps 0;
```

```
Qy    658  ggcggcaagctggaacagaagcgttacttcaagcctcagttcccagtacagaaggtcgta  717
               |-|||    ||-||||| ||     |||||||    ||| |||   || ||-|
Db   1389  GGCAGGCTGGAAGAGCAGAAGGGACCCGACGTCATGGCGGCGGCCATCCGCCAGCTCATG 1448

Qy    718  aaggtaaggacgaggacctcttcgatcgcattgccaagtgttggaggatagcgtcgaa  777
               |  |-| |||   |||||  |    |||| ||      ||||| ||
Db   1449  GAGATGGTGGAGGACGTCCAGATCGTTCTGCTGGCACGGGCAAGAAGAGTTCGAGCGC 1508

Qy    778  aagcatatgctgtccgacgtgaccgtaggctcgtccttcgggcggactgactcaacc  837
               |||||  ||  ||  ||| |   |   |  |-|  | | ||
Db   1509  ATGCTCATGAGCGCCGAGGAGAGTTCCAGGCAAGGTGCGCCGCCGTCACCTCACCACCGGTTCAAC 1558

Qy    838  gcaattgcgcgcttgcaaagcgccacaaccctgacctgctcacttcaccacccggtttc  897
               || |||     ||||  ||||| |  |||||| ||||||
Db   1569  GGGGCGCTGGCCACCACCATGCCGGCCGCGAGCGTGCTCGCCGTCACCGTCCACCAGCCGCCTTC 1628

Qy    898  gagcgtgaaggtactcggaggtcgatgtg  927
               |||||   ||      |-| |||
Db   1629  GAGCCCTGCGGCCTCATCCAGCTGCAGGGG 1658
```

RESULT 12
US-08-997-080-92
; Sequence 92, Application US/08997080
; Patent No. 5968524
; GENERAL INFORMATION:
; APPLICANT: WATSON, JAMES D.
; APPLICANT: TAN, PAUL L.J.
; TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-
; NUMBER OF SEQUENCES: 194
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Law Offices of Ann W. Speckman
;   STREET: 2601 Elliott Avenue, Suite 4185
;   CITY: Seattle
;   STATE: WA
;   COUNTRY: USA
;   ZIP: 98121
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/997,080
;   FILING DATE:
;   CLASSIFICATION:
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Sleath, Janet
;   REGISTRATION NUMBER: 37,007
;   REFERENCE/DOCKET NUMBER: 11000.1007
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 206-269-0565
;   TELEFAX: 206-269-0563
;   TELEX:
; INFORMATION FOR SEQ ID NO: 92:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 323 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-08-997-080-92

Query Match   1.8%;   Score 34.4;   DB 2;   Length 323;
Best Local Similarity 53.8%;   Pred. No. 0.8;
Matches 71;   Conservative 0;   Mismatches 61;   Indels 0;   Gaps 0;

Qy   446 agccactgttctacgcaccaccgagcatggccacgtgttcctcagagaagagacca 505
               || |||||||| |||||||| | ||| |||||| ||| |||||| |||-||||
```

Db    191 AGGACCTGTTGGACTCGAGCATCGACGAGGGCCCGCAAGTTCACCGGCGCCGTACATGACCG 250

Qy    506 tcttggagatggccgaggagatgaatctagatctgggccttgataagcgcaccattgagc 565

Db    251 GCATGGTCGGTCGTCTCGCCTACACAAGGCAGCCACCGGACGGCGGATATCCGCACCATCGACG 310

Qy    566 actacgtggacc 577

Db    311 ACCTCTGGGATC 322

RESULT 13
US-08-997-362-92
; Sequence 92, Application US/08997362
; Patent No. 5985287
; GENERAL INFORMATION:
;   APPLICANT: Tan, Paul
;   APPLICANT: Hiyama, Jun
;   APPLICANT: Visser, Elizabeth
;   APPLICANT: Skinner, Margot
;   APPLICANT: Scott, Linda
;   APPLICANT: Prestidge, Ross
;   TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
;   TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
;   NUMBER OF SEQUENCES: 194
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Law Offices of Ann W. Speckman
;     STREET: 2601 Elliott Avenue, Suite 4185
;     CITY: Seattle
;     STATE: WA
;     COUNTRY: USA
;     ZIP: 98121
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette
;     COMPUTER: IBM Compatible
;     OPERATING SYSTEM: DOS
;     SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/997,362
;     FILING DATE:
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970
;     FILING DATE: June 12, 1997
;     APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347
;     FILING DATE: August 29, 1996
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Sleath, Janet
;     REGISTRATION NUMBER: 37,007
;     REFERENCE/DOCKET NUMBER: 11000.1002c2
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 206-269-0565
;     TELEFAX: 206-269-0563
;     TELEX:
; INFORMATION FOR SEQ ID NO: 92:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 323 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: Genomic DNA
US-08-997-362-92

Query Match        1.8%;    Score 34.4;  DB 2;  Length 323;
Best Local Similarity  53.8%;  Pred. No. 0.8;
Matches   71;  Conservative   0;  Mismatches   61;  Indels   0;  Gaps   0;

Qy    446 agccactgttctacgcaaccaccgagcatggcaccgtgttctcctcagagagaagagacca 505
          || |||||| ||| ||| || ||| || | ||| || || ||| || | |||||| ||||
Db    191 AGGACCTGTTGGACTCGAGCATCGACGAGGGCCCGCAAGTTCACCGGCGCCGTACATGACCG 250

Qy    506 tcttggagatggccgaggagatgaatctagatctgggccttgataagcgcaccattgagc 565

RESULT 14
US-08-873-970-92
; Sequence 92, Application US/08873970
; Patent No. 6001361
; GENERAL INFORMATION:
;   APPLICANT: Tan, Paul
;   APPLICANT: Hiyama, Jun
;   APPLICANT: Visser, Elizabeth
;   APPLICANT: Skinner, Margot
;   APPLICANT: Scott, Linda
;   APPLICANT: Prestidge, Ross
;   TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
;   TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
;   NUMBER OF SEQUENCES: 106
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Law Offices of Ann W. Speckman
;     STREET: 2601 Elliott Avenue, Suite 4185
;     CITY: Seattle
;     STATE: WA
;     COUNTRY: USA
;     ZIP: 98121
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette
;     COMPUTER: IBM Compatible
;     OPERATING SYSTEM: DOS
;     SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/873,970
;     FILING DATE:
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/705,347
;     FILING DATE: 29-AUG-1996
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Sleath, Janet
;     REGISTRATION NUMBER: 37,007
;     REFERENCE/DOCKET NUMBER: 11000.1002C1
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 206-269-0565
;     TELEFAX: 206-269-0563
;     TELEX:
; INFORMATION FOR SEQ ID NO: 92:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 323 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: Genomic DNA
US-08-873-970-92

Query Match        1.8%;    Score 34.4;  DB 3;  Length 323;
Best Local Similarity  53.8%;  Pred. No. 0.8;
Matches   71;  Conservative   0;  Mismatches   61;  Indels   0;  Gaps   0;

Qy    446 agccactgttctacgcaaccaccgagcatggcaccgtgttctcctcagagagaagagacca 505
          || |||||| ||| ||| || ||| || | ||| || || ||| || | |||||| ||||
Db    191 AGGACCTGTTGGACTCGAGCATCGACGAGGGCCCGCAAGTTCACCGGCGCCGTATACGACCG 250

Qy    506 tcttggagatggccgaggagatgaatctagatctgggccttgataagcgcaccattgagc 565
          || |||| ||| || |||||||| || || || ||||| || | ||||| | |||||||
Db    251 GCATGGTCGGTCGTCTCGCCTACACAAGGCAGCCACCGGACGGCGGATATCCGCACCATCGACG 310

Qy    566 actacgtggacc 577
          || | ||| |

Db    311 ACCTCTGGGATC 322

Search completed: September 20, 2002, 07:14:05
Job time: 4797 sec

RESULT 15
US-09-095-855-92
; Sequence 92, Application US/09095855
; Patent No. 6160093
; GENERAL INFORMATION:
;  APPLICANT: Tan, Paul
;  APPLICANT: Visser, Elizabeth
;  APPLICANT: Skinner, Margot
;  APPLICANT: Prestidge, Ross
;  TITLE OF INVENTION: Compounds and Methods for
;  TITLE OF INVENTION: Treatment and Diagnosis of Mycobacterial Infections
;  NUMBER OF SEQUENCES: 208
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Law Offices of Ann W. Speckman
;   STREET: 2601 Elliott Avenue, Suite 4185
;   CITY: Seattle
;   STATE: WA
;   COUNTRY: USA
;   ZIP: 98121
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/095,855
;   FILING DATE:
;   CLASSIFICATION:
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/705,347
;   FILING DATE: 29-AUG-1996
;   APPLICATION NUMBER: 08/873,970
;   FILING DATE: 12-JUN-1997
;   APPLICATION NUMBER: 08/997,362
;   FILING DATE: 23-DEC-1997
;  ATTORNEY/AGENT INFORMATION:
;   NAME: Sleath, Janet
;   REGISTRATION NUMBER: 37,007
;   REFERENCE/DOCKET NUMBER: 11000.1002c3
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 206-269-0565
;   TELEFAX: 206-269-0563
;   TELEX:
;  INFORMATION FOR SEQ ID NO: 92:
;   SEQUENCE CHARACTERISTICS:
;    LENGTH: 323 base pairs
;    TYPE: nucleic acid
;    STRANDEDNESS: single
;    TOPOLOGY: linear
;   MOLECULE TYPE: Genomic DNA
US-09-095-855-92

Query Match             1.8%;  Score 34.4;  DB 4;  Length 323;
Best Local Similarity  53.8%;  Pred. No. 0.8;
Matches  71;  Conservative  0;  Mismatches  61;  Indels  0;  Gaps  0;

Qy   446 agccactgttctacgcaaccaccgagcatggcaccgtgttcctccctcagagaagaagacca 505
             ||  ||||| || ||  |||| ||   |  ||| | | |   ||||| |||  |||||
Db   191 AGGACCTGTTGGACTCGAGCATCGACGAGGCGCCGCAAGTTCACCGCGCCCGTACATGACCG 250

Qy   506 tcttggagatggccgaggagatgaatctagatctggccttgataagcgcaccattgagc 565
             |  ||| ||  |||| || |  |  ||   ||| |||| ||||||| ||
Db   251 GCATGGTCGGTCGCCTACAACAGGCAGCCACCGGACGCGATATCCGACCACCATCGACG 310

Qy   566 actacgtggacc 577
             || ||||| |||
Db   311 ACCTCTGGGATC 322

us-09-786-474-1.rni

```
                 Copyright  GenCore version 4.5
                 Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:          September 20, 2002, 04:48:28 ; Search time 1796.7 Seconds
                 (without alignments)
                 14423.195 Million cell updates/sec

Title:           US-09-786-474-1
Perfect score:   1920
Sequence:        1 atgtgcggcccttcttggcat.........gctcctaccggtcgagctt 1920

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

arched:          13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters:          27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     EST:*
               1:  em_estba:*
               2:  em_esthum:*
               3:  em_estin:*
               4:  em_estmu:*
               5:  em_estov:*
               6:  em_estpl:*
               7:  em_estro:*
               8:  em_htc:*
               9:  gb_est1:*
              10:  gb_est2:*
              11:  gb_htc:*
              12:  gb_gss:*
              13:  em_gss_hum:*
              14:  em_gss_inv:*
              15:  em_gss_pln:*
              16:  em_gss_vrt:*
```

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c						
1	58.6	3.1	635	12	BH001105	BH001105 G179 Gemm
2	56.2	2.9	552	9	AA697638	AA697638 HLO2916.5
3	56.2	2.9	565	9	AI516024	AI516024 LD41758.5
4	56.2	2.9	575	9	AI387646	AI387646 GH18225.5
5	56.2	2.9	601	10	BI630083	BI630083 RH59048.5
6	56.2	2.9	602	10	BI239433	BI239433 RE35917.5
7	56.2	2.9	609	9	AI062991	AI062991 GH02388.5
8	56.2	2.9	616	10	BI641347	BI641347 SD24579.5
9	56.2	2.9	626	9	BI212840	BI212840 RE18085.5
10	56.2	2.9	630	9	AI389247	AI389247 GH20374.5
11	56.2	2.9	644	10	BI635460	BI635460 SD16825.5
12	56.2	2.9	693	10	BI142084	BI142084 SD16082.5
13	56.2	2.9	701	10	BI641950	BI641950 SD25636.5
14	55	2.9	640	10	BI719579	BI719579 1031044E0
15	54.6	2.8	620	10	BI354992	BI354992 GM29151.5
16	53.6	2.8	601	12	BI527119	BI527119 1024078D0
17	53.6	2.8	664	10	BI874100	BI874100 963113D09
18	51.4	2.7	560	9	AI064256	AI064256 GH04615.5
19	48.8	2.5	404	10	BG690202	BG690202 338655 BA
20	47.2	2.5	503	9	AV433013	AV433013 AV433013
21	45.6	2.4	408	10	BG015755	BG015755 CMI-GN032
22	45.6	2.4	418	10	BI261522	BI261522 602953585
23	45.6	2.4	418	10	BI261879	BI261879 602953839
24	45.6	2.4	506	10	BG423223	BG423223 602450541
25	45.6	2.4	544	10	BG479329	BG479329 602526358
26	45.6	2.4	620	9	BE207639	BE207639 bb64b11.y
27	45.6	2.4	654	10	BE262677	BE262677 601150869
28	45.6	2.4	665	10	BE268144	BE268144 601125194
29	45.6	2.4	680	10	BM042280	BM042280 603616356
30	45.6	2.4	684	10	BM051714	BM051714 603638511
31	45.6	2.4	693	10	BE901630	BE901630 601677773
32	45.6	2.4	703	10	BI087480	BI087480 602851929
33	45.6	2.4	740	10	BE735080	BE735080 601569488
34	45.6	2.4	757	10	BI225629	BI225629 602948973
35	45.6	2.4	815	10	BG683088	BG683088 602651765
36	45.6	2.4	826	10	BI755348	BI755348 603024923
37	45.6	2.4	830	10	BI562572	BI562572 603256505
38	45.6	2.4	830	10	BE396356	BE396356 601288712
39	45.6	2.4	865	10	BI826065	BI826065 603076485
40	45.6	2.4	881	10	BE901446	BE901446 601674791
41	45.6	2.4	906	10	BM461856	BM461856 AGENCOURT
42	45.6	2.4	918	9	AL557155	AL557155 AL557155
43	45.6	2.4	944	10	BG420309	BG420309 602448328
44	45.6	2.4	945	10	BE270764	BE270764 600943910
45	45.6	2.4	952	10	BE560921	BE560921 601346355

ALIGNMENTS

```
RESULT 1
BH001105/c
LOCUS       BH001105              635 bp    DNA     linear   GSS 01-JUL-2001
DEFINITION  G179 Gemmata obscuriglobus phagemid library Gemmata obscuriglobus
            genomic clone G179 similar to putative asparagine synthetase
            (glutamine hydrolyzing), DNA sequence.
ACCESSION   BH001105
VERSION     BH001105.1  GI:14579871
KEYWORDS    GSS.
SOURCE      Gemmata obscuriglobus.
  ORGANISM  Gemmata obscuriglobus
            Bacteria; Planctomycetes; Planctomycetales; Planctomycetaceae; Gemmata.
REFERENCE   1  (bases 1 to 635)
  AUTHORS   Jenkins,C., Kedar,V. and Fuerst,J.A.
  TITLE     Gene discovery from sequence tags generated using genomic DNA
            libraries constructed from representatives of the planctomycete
            division of the Domain Bacteria
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Fuerst JA
            Department of Microbiology and Parasitology
            University of Queensland
            Brisbane, QLD 4072, Australia
            Tel: +617 3365 4643
            Fax: +617 3365 4620
            Email: fuerst@bioscl.uq.edu.au
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..635
                     /organism="Gemmata obscuriglobus"
                     /strain="ACM 2246"
                     /db_xref="taxon:114"
                     /clone="G179"
                     /clone_lib="Gemmata obscuriglobus phagemid library"
                     /lab_host="Escherichia coli strain XL1-Blue"
                     /note="vector: pBluescript II SK(-) phagemid; Site_1:
                     BamHI; Site_2: BamHI; Purified genomic DNA from Gemmata
                     obscuriglobus was restricted with Sau3AI to give fragments
                     of 400-3000bp and ligated into the BamHI site of the
                     pBluescript II SK(-) phagemid. The ligated DNA was
                     electroporated into E. coli XL1-Blue."
```

Right column (Result block):

```
/sex="male and female"
/dev_stage="adult"
/lab_host="SOLR"
/note="Organ: head--brain & sensory organ; Vector:
BlueScript SK; Site_1: EcoRI; Site_2: XhoI; Constructed
using Stratagene ZAP-cDNA Synthesis kit. Oligo dT-primed
and directionally cloned at EcoRI and XhoI in BlueScript
SK(+/-)"

BASE COUNT     133 a    122 c    157 g    140 t
ORIGIN

Query Match           2.9%;   Score 56.2; DB 9;  Length 552;
Best Local Similarity 53.4%;  Pred. No. 0.00047;
Matches 118; Conservative    0; Mismatches 103; Indels  0; Gaps  0:

Qy   234 caacggtgagatctacaactacgttgagctgcgtaaagagcctcggatttgggatatac 293
         ||||||||| |||||||| | |||||||||||||||  | ||| ||||  |||||||
Db   119 CAATGGAGAGATCTACAATTATCTGGAGCTTTCGGCTGGAGATTGCGAAAAAGCCTGGCTC 178

Qy   294 ctttaatacttcggcgatggcgagccaattgtgttcgtttccaccactggggcgagtc 353
         ||| ||||||||||||||| ||||| |||||||| ||  ||  ||||||||||  || |
Db   179 CTACAACCCCATGAGCATTGGCACGTGATACTGGAACTGTACCAGGATTATGGAAAGGA 238

Qy   354 cgtggtcgagcatctccgcggaatgttcggcattgcattggatacaaaggaaagtc 413
         |||| |||||||| | ||  ||  ||| ||||||||||| ||||||||||
Db   239 TCTTCTGCAATACATCACTGGAATGTTTGCTTTGCCTTGTACGATAGGAAGACCAAGGA 298

Qy   414 gctttcctggcgtgatcagttcggcatcaagccactgt 454
         ||||||||  |||| | |||||||||||||| ||||||
Db   299 AGTGCTTCTTGCCCGAGACCCCTTTGGCATTATACCCATGT 339

RESULT   3
AI516024
LOCUS          AI516024                 565 bp    mRNA   linear EST 19-APR-2001
DEFINITION     LD41758.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
               melanogaster cDNA clone LD41758 5prime, mRNA sequence.
ACCESSION      AI516024
VERSION        AI516024.1    GI:4419124
KEYWORDS       EST.
SOURCE         fruit fly.
ORGANISM       Drosophila melanogaster
               Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
               Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
               Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 565)
AUTHORS        Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
               Lewis,S. and Rubin,G.M.
TITLE          BDGP/HHMI Drosophila EST Project
JOURNAL        Unpublished (2001)
COMMENT        Contact: Stapleton, M.
               BDGP
               Lawrence Berkeley National Lab
               One Cyclotron Rd, Berkeley, CA 94720, USA
               Fax: 510 486 6798
               Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
               Plate: 417  row: E  column: 10
               High quality sequence stop: 529.
FEATURES       Location/Qualifiers
     source    1..565
               /organism="Drosophila melanogaster"
               /db_xref="taxon:7227"
               /clone="LD41758"
               /clone_lib="LD Drosophila melanogaster embryo pOT2"
               /sex="male and female"
               /dev_stage="0 to 24 hours mixed stage embryonic"
               /lab_host="XL1 Blue"
               /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
               XhoI; Sized fractionated cDNAs were directly ligated into
               pOT2."
BASE COUNT     142 a    125 c    162 g    136 t
ORIGIN
```

Left column:

```
BASE COUNT     103 a    223 c    197 g    112 t
ORIGIN

Query Match           3.1%;   Score 58.6; DB 12;  Length 635;
Best Local Similarity 50.6%;  Pred. No. 0.00012;
Matches 220; Conservative    0; Mismatches 209; Indels  6; Gaps  3;

Qy   725 aggagaggacctcttcgatcgcattgccaggtgttggaggatagcgtcgaaaagcata 784
         | ||||||||| | || |||||| | ||||||  ||||||| ||| ||| |||
Db   434 ATGAGAAGAAGGTCGTCGACGGGTTCGAGGAGGTGCTGTACG-CCCGGTCCAGCGCAGG 376

Qy   785 tgcgtgccgacgtgaccgtaggctgttcctttccggcgcgcattgactcaaccgcaattg 844
         | ||||||||||||| |||||||| |||  ||  |||| |||  |||| ||| |
Db   375 GTGTGGCCGACGTGCCCGTAGTGTCTACCTGAGCCGCGGCGGCGGTGGACTCGAGCGTCG 316

Qy   845 cgccgcttgcaaagc--gccacaaccctgacctgctcaacttcaccaccacggtttcgagcg 902
         ||||||| ||||||   | | | ||| |||||||||| ||||||||  ||  |
Db   315 TGGCCATGGCGAACAAGGCGATGGGCGCGCGCCCATCCGACGTTCACCGTCTCGGTCAACGC 256

Qy   903 tgaaggctactcggaggtcgatgtggctgcgcgagtccgcgctgcgattggcgctgagca 962
         || |||||||||||||| | |||  |||| |||||||  | | ||||| ||  ||||||
Db   255 GAAAGGGCTGCACGACAGTCCGAGCGCCGCTGGCCAAGCACCTGGGCTGTTCGCC 196

Qy   963 catcgtgaagattgtctcgcctgaggaatacgccaacgcgattcctaagatcatgtggta 1022
         ||  | ||||  |  || ||| ||  |||  | ||||||  | | ||||||||| |
Db   195 GGTAGTAGTGCACTGCGAGTCCGGGCGCACGACCAGCTTCCGCGGCGCTACCCGGCTGATCGCCGC 136

Qy   1023 cttggatgatcctggctagctgacccatcat---tggtcccgctgtacttcgttgcagcgga 1079
          |||| |  || ||  |||   ||||||  |    ||||||| ||||| | |||||| ||||
Db   135 CGCCGAGTTCCCGGTCATCGACACACGTCCCTGCCTGCCCTGCCCTCCTTCATTTAGCCGGAGCGGT 76

Qy   1080 agacgtaagcacgtcaaggttgtgctgtctcgcgaggggcgcagatgagctgttcggtgg 1139
          ||||||| ||||  |||||||||||||||| | ||||||||||| ||||||||| |||
Db   75  GCATCCAAACGGGTACAAGGTGCGCTCGCTCCTAACTGGTGAGGGGCCCGACGAGTGGCTGGGCGG 16

Qy   1140 atacaccattacaa 1154
          || | | | ||||
Db   15  GTACTCGTGGTTCAA 1

RESULT   2
AA697638
LOCUS          AA697638                 552 bp    mRNA   linear EST 19-APR-2001
DEFINITION     HL02916.5prime HL Drosophila melanogaster head BlueScript
               Drosophila melanogaster cDNA clone HL02916 5prime, mRNA sequence.
ACCESSION      AA697638
VERSION        AA697638.1    GI:2700567
KEYWORDS       EST.
SOURCE         fruit fly.
ORGANISM       Drosophila melanogaster
               Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
               Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
               Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 552)
AUTHORS        Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
               Lewis,S. and Rubin,G.M.
TITLE          BDGP/HHMI Drosophila EST Project
JOURNAL        Unpublished (2001)
COMMENT        Contact: Stapleton, M.
               BDGP
               Lawrence Berkeley National Lab
               One Cyclotron Rd, Berkeley, CA 94720, USA
               Fax: 510 486 6798
               Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
               Plate: 29  row: B  column: 4
               High quality sequence stop: 505.
FEATURES       Location/Qualifiers
     source    1..552
               /organism="Drosophila melanogaster"
               /db_xref="taxon:7227"
               /clone="HL02916"
               /clone_lib="HL Drosophila melanogaster head BlueScript"
```

Qy   294 cttaatacttctggcgatggcgagccaattgttgtcggtttccaccactgggcgagtc 353
           ||| || |||| ||||| ||| ||| || || |||||||||||||||||| |||||
Db   381 CTACAACCCCATGAGCGATTGCCACGTGATACTGGAACTGGAACTGTACCAGGATTATGGAAAGGA 440

Qy   354 cgtggtcgagcatctccgcggaatggtcggcattggcattggga tacaaagaaaaagtc 413
           | | ||||||||||||||||||||||||||||||| ||||| ||| ||| |||
Db   441 TCTTCTGCAATACATCACTGGAATGTTTGCTTTTGCCTTGCTTTGCCTTGTACGATAGGAAGACCAAGGA 500

Qy   414 gctttcctgcgcgtgatcagttcggcatcaagccactgt 454
           | ||||||||||||| ||||||||||||||||||| |||
Db   501 AGTGCTTCTTGCCCGAGACCCCTTTGGCATTATACCCATGT 541

RESULT   5
BI630083
LOCUS   BI630083   601 bp   mRNA   linear   EST 10-SEP-2001
DEFINITION   RH59048.5prime RH Drosophila melanogaster normalized Head pFlc-1
             Drosophila melanogaster cDNA clone RH59048 5, mRNA sequence.
ACCESSION   BI630083
VERSION   BI630083.1 GI:15532293
KEYWORDS   EST.
SOURCE   fruit fly.
ORGANISM   Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
             Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
             Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 601)
AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
             ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
             ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
             Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
             Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
             ,G.M.
TITLE   BDGP/HHMI RH Drosophila EST Project
JOURNAL   Unpublished (2001)
COMMENT   Contact: Stapleton, M.
             BDGP
             Lawrence Berkeley National Lab
             One Cyclotron Rd, Berkeley, CA 94720, USA
             Fax: 510 486 6798
             Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
             hit genomic AE003593: arm:3L [20799555, 21112669]
             estimated-cyto:78A2-78C7: 08/23/2001
             Plate: RH.590 row: D column: 12
             High quality sequence stop: 578.
FEATURES   Location/Qualifiers
     source   1..601
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="RH59048"
                     /clone_lib="RH Drosophila melanogaster normalized Head
                     pFlc-1"
                     /sex="male and female"
                     /dev_stage="Adult"
                     /lab_host="DH5-alpha TonA"
                     /note="Organ: head; Vector: pFlc1; Site_1: xhoI; Site_2:
                     BamHI; Library was kindly generated by Piero Carnicni at
                     the RIKEN. The library was normalized and excised using
                     Cre recombinase. Plasmid cDNA library."
BASE COUNT   157 a   130 c   164 g   150 t
ORIGIN

Query Match   2.9%; Score 56.2; DB 10; Length 601;
Best Local Similarity 53.4%; Pred. No. 0.00049;
Matches 118; Conservative   0; Mismatches 103; Indels   0; Gaps   0;

Qy   234 caacggtgagatctacaactacgttgagctgcgtaaagagctctcggatttgggatatac 293
           |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db   341 CAATGGAGAGATCTACAATTATCTGGAGCTTCGGCTGGAGATTGCGAAAAAGCCTGGCTC 400

Qy   294 cttaatacttctggcgatggcgagccaattgttgtcggtttccaccactgggcgagtc 353
           ||| || |||| ||||| ||| ||| ||

```
Db   401 CTACAACCCCATGAGCGATTGCCACGTGATACTGGAACTGTACCAGGATTATGGAAAGGA 460

Qy   354 cgtggtcgagcatctccgcggaatgttcggcattgccatttgggatacaaaggaaagtc 413
             |||||| | ||||||||||||||||||||||||||||||| |||  |||
Db   461 TCTTCTGCAATACATCACTGGAATGTTTGCTTTGCCTTGTACGATAGGAAGACCAAGGA 520

Qy   414 gctttcctgcgcgtgatcagttcggcatcaagccactgt 454
             | |||||| || |||||| ||  |  ||||||||
Db   521 AGTGCTTCTTGCCCGAGACCCCTTTGGCATTATACCCATGT 561


RESULT 6
BI239433
LOCUS       BI239433                602 bp    mRNA    linear   EST 12-JUL-2001
DEFINITION  RE35917.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
            Drosophila melanogaster cDNA clone RE35917 5, mRNA sequence.
ACCESSION   BI239433
VERSION     BI239433.1  GI:14707974
KEYWORDS    EST.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 602)
AUTHORS     Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
            ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
            ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
            Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
            ,G.M.
TITLE       BDGP/HHMI RE Drosophila EST Project
JOURNAL     Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003593::arm:3L [20799555,21112669]
            estimated-cyto:78A2-78C7: 05/12/2001
            Plate: RE.359  row: B  column: 5
            High quality sequence stop: 508.
FEATURES             Location/Qualifiers
     source          1..602
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="RE35917"
                     /clone_lib="RE Drosophila melanogaster normalized Embryo
                     pFlc-1"
                     /sex="male and female"
                     /dev_stage="0-24 hours mixed stage embryonic"
                     /lab_host="DH5-alpha TonA"
                     /note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
                     BamHI; Library was kindly generated by Piero Carninci at
                     the RIKEN. The library was normalized and excised using
                     Cre recombinase. Plasmid cDNA library."
BASE COUNT    158 a    129 c    166 g    149 t
ORIGIN

Query Match     2.9%;   Score 56.2;  DB 10;  Length 602;
Best Local Similarity  53.4%;   Pred. No. 0.00049;
Matches 118;  Conservative 0;  Mismatches 103;  Indels 0;  Gaps 0;

Qy   234 caacggtgagatctacaactacgttgagctgcgtaaagagctctcggatttgggatatac 293
             ||||||||| || | |||||||| ||  ||||||| |||||||
Db   340 CAATGGAGAGATCTACAATTATCTGGAGCTTCGGGATTGCGAAAAAGCCGTGGCTC 399

Qy   294 ctttaatacttctggcgatggcgagccaattgttgtcggtttccaccactggggcgagtc 353
             |||||||||||||||||||| |||| |||| |||| || ||||||
Db   400 CTACAACCCCATGAGCGATTGCCACGTGATACTGGAACTGTACCAGGATTATGGAAAGGA 459
```

```
Qy   354 cgtggtcgagcatctccgcggaatgttcggcattgccatttgggatacaaaggaaagtc 413
             |||||| | |||||||||||||||||||||||||||||||||  |||
Db   460 TCTTCTGCAATACATCACTGGAATGTTTGCTTTGCCTTGTACGATAGGAAGACCAAGGA 519

Qy   414 gctttcctgcgcgtgatcagttcggcatcaagccactgt 454
             | |||||| || |||||| ||  |  ||||||||
Db   520 AGTGCTTCTTGCCCGAGACCCCATTTGGCATTATACCCATGT 560


RESULT 7
AI062991
LOCUS       AI062991                609 bp    mRNA    linear   EST 19-APR-2001
DEFINITION  GH02388.5prime GH Drosophila melanogaster head pOT2 Drosophila
            melanogaster cDNA clone GH02388 5prime, mRNA sequence.
ACCESSION   AI062991
VERSION     AI062991.1  GI:3338830
KEYWORDS    EST.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 609)
AUTHORS     Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
TITLE       BDGP/HHMI Drosophila EST Project
JOURNAL     Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 23  row: H  column: 4
            High quality sequence stop: 564.
FEATURES             Location/Qualifiers
     source          1..609
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="GH02388"
                     /clone_lib="GH Drosophila melanogaster head pOT2"
                     /sex="male and female"
                     /dev_stage="adult"
                     /lab_host="DH5 - alpha"
                     /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2. Plasmid cDNA library."
BASE COUNT    146 a    145 c    167 g    151 t
ORIGIN

Query Match     2.9%;   Score 56.2;  DB 9;  Length 609;
Best Local Similarity  53.4%;   Pred. No. 0.0005;
Matches 118;  Conservative 0;  Mismatches 103;  Indels 0;  Gaps 0;

Qy   234 caacggtgagatctacaactacgttgagctgcgtaaagagctctcggatttgggatatac 293
             ||||||||| || | |||||||| ||  ||||||| |||||||
Db    12 CAATGGAGAGATCTACAATTATCTGGAGCTTCGGGATTGCGAAAAAGCCGTGGCTC 71

Qy   294 ctttaatacttctggcgatggcgagccaattgttgtcggtttccaccactggggcgagtc 353
             |||||||||||||||||||| |||| |||| |||| || ||||||
Db    72 CTACAACCCCATGAGCGATTGCCACGTGATACTGGAATGTTGGGATTATGGAAAGGA 131

Qy   354 cgtggtcgagcatctccgcggaatgttcggcattgccatttgggatacaaaggaaagtc 413
             |||||| | |||||||||||||||||||||||||||||||||  |||
Db   132 TCTTCTGCAATACATCACTGGAATGTTTGCTTTGCCTTGTACGATAGGAAGACCAAGGA 413

Qy   414 gctttcctgcgcgtgatcagttcggcatcaagccactgt 454
             | |||||| || |||||| ||  |  ||||||||
Db   414 AGTGCTTCTTGCCCGAGACCCCTTTGGCATTATACCCATGT 454

Db   192 AGTGCTTCTTGCCCGAGACCCCTTTGGCATTATACCCATGT 232


RESULT 8
```

BI641347
LOCUS   BI641347   616 bp   mRNA   linear   EST 10-SEP-2001
DEFINITION SD24579.5prime SD Drosophila melanogaster Schneider L2 cell culture
            pOT2 Drosophila melanogaster cDNA clone SD24579 5, mRNA sequence.
ACCESSION   BI641347
VERSION   BI641347.1 GI:15543557
KEYWORDS   EST.
SOURCE   fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 616)
AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su.,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
TITLE   BDGP/HHMI Drosophila EST Project
JOURNAL   Unpublished (2001)
MMENT   Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003593: arm:3L [20799555,21112669]
            estimated-cyto:78A2-78C7: 05/19/2001
            Plate: SD.245 row: G column: 7
            High quality sequence stop: 511.
FEATURES   Location/Qualifiers
     source 1..616
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone="SD24579"
            /clone_lib="SD Drosophila melanogaster Schneider L2 cell
            culture pOT2"
            /lab_host="DH5-alpha"
            /note="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
            fractionated cDNAs were directly ligated into pOT2.
            Plasmid cDNA library."
BASE COUNT   161 a   133 c   173 g   149 t
ORIGIN

Query Match   2.9%;   Score 56.2; DB 10; Length 616;
Best Local Similarity   53.4%;   Pred. No. 0.0005;
Matches 118; Conservative   0; Mismatches 103; Indels   0; Gaps   0;

Qy   234 caacggtgagatctacaactacgttgagctgcgtaaagagctctcggatttgggatatac 293
           |||-|||||||||||||||||| || |||||||||||||||||||||||||||
Db   316 CAATGGAGAGAGATCTACAATTATCTGGAGCTTCGGCTTGTGCGGAAAAAGCCTGGCTC 375

Qy   294 ctttaatacttctggcgatggcggagccaattgttgtcggttccaccactgggcgcgagtc 353
           || || |||||||||||||||||||| |||| || |||||||||||||
Db   376 CTACAACCCCATGAGCGATTGCCACGTGATACTGAACGTACCAGGATTATGTGGAAAGGA 435

Qy   354 cgtggtcgagcatctccgcggaatgttcggcattgccatttgggatacaaaggaaaagtc 413
           |-|||||||| | |||||| || ||||| |||| |||| ||| |||
Db   436 TCTTCTGCAATACATCACTGGAATGTTTGCCTTTGCTTTGCCTGTGTACGATAGGAAGACCAAGA 495

Qy   414 gctttcctgcgcgtgatcagttcggcatcaagccactgt 454
           | ||||||||| ||| || |||| ||||||||||||
Db   496 AGTGCTCTTGCCCGAGACCCCTTTGGCATTATACCCATGT 536

RESULT 9
BI212840
LOCUS   BI212840   626 bp   mRNA   linear   EST 11-JUL-2001
DEFINITION RE18085.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
            Drosophila melanogaster cDNA clone RE18085 5, mRNA sequence.
ACCESSION   BI212840
VERSION   BI212840.1 GI:14690555
KEYWORDS   EST.
SOURCE   fruit fly.
ORGANISM   Drosophila melanogaster

REFERENCE   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
            1 (bases 1 to 626)
AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
            ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
            ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
            Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
            ,G.M.
TITLE   BDGP/HHMI RE Drosophila EST Project
JOURNAL   Unpublished (2001)
COMMENT   Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003593: arm:3L [20799555,21112669]
            estimated-cyto:78A2-78C7: 04/12/2001
            Plate: RE.180 row: H column: 1
            High quality sequence stop: 592.
FEATURES   Location/Qualifiers
     source 1..626
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone="RE18085"
            /clone_lib="RE Drosophila melanogaster normalized Embryo
            pFlc-1"
            /sex="male and female"
            /dev_stage="0-24 hours mixed stage embryonic"
            /lab_host="DH5-alpha TonA"
            /note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
            BamHI; Library was kindly generated by Piero Carninci at
            the RIKEN. The library was normalized and excised using
            Cre recombinase. Plasmid cDNA library."
BASE COUNT   162 a   134 c   175 g   155 t
ORIGIN

Query Match   2.9%;   Score 56.2; DB 10; Length 626;
Best Local Similarity   53.4%;   Pred. No. 0.0005;
Matches 118; Conservative   0; Mismatches 103; Indels   0; Gaps   0;

Qy   234 caacggtgagatctacaactacgttgagctgcgtaaagagctctcggatttgggatatac 293
           |||-||||||||||||||||| || |||||||||||||||||||||||||||
Db   340 CAATGGAGAGAGATCTACAATTATCTGGAGCTTCGGCTTGTGCGGAAAAAGCCTGGCTC 399

Qy   294 ctttaatacttctggcgatggcggagccaattgttgtcggttccaccactgggcgcgagtc 353
           || || |||||||||||||||||||| |||| || |||||||||||||
Db   400 CTACAACCCCATGAGCGATTGCCACGTGATACTGAACGTACCAGGATTATGTGGAAAGGA 459

Qy   354 cgtggtcgagcatctccgcggaatgttcggcattgccatttgggatacaaaggaaaagtc 413
           |-|||||||| | |||||| || ||||| |||| |||| ||| |||
Db   460 TCTTCTGCAATACATCACTGGAATGTTTGCCTTTGCTTTGCCTGTGTACGATAGGAAGACCAAGA 519

Qy   414 gctttcctgcgcgtgatcagttcggcatcaagccactgt 454
           | ||||||||| ||| || |||| ||||||||||||
Db   520 AGTGCTCTTGCCCGAGACCCCTTTGGCATTATACCCATGT 560

RESULT 10
AI389247
LOCUS   AI389247   630 bp   mRNA   linear   EST 19-APR-2001
DEFINITION GH20374.5prime GH Drosophila melanogaster head pOT2 Drosophila
            melanogaster cDNA clone GH20374 5prime, mRNA sequence.
ACCESSION   AI389247
VERSION   AI389247.1 GI:4203258
KEYWORDS   EST.
SOURCE   fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

```
REFERENCE              Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
  AUTHORS              1 (bases 1 to 630)
                       Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
                       Lewis,S. and Rubin,G.M.
  TITLE                BDGP/HHMI Drosophila EST Project
  JOURNAL              Unpublished (2001)
COMMENT                Contact: Stapleton, M.
                       BDGP
                       Lawrence Berkeley National Lab
                       One Cyclotron Rd, Berkeley, CA 94720, USA
                       Fax: 510 486 6798
                       Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
                       Plate: 203 row: G column: 2
                       High quality sequence stop: 408.
FEATURES               Location/Qualifiers
  source               1..630
                       /organism="Drosophila melanogaster"
                       /db_xref="taxon:7227"
                       /clone="GH20374"
                       /clone_lib="GH Drosophila melanogaster head pOT2"
                       /sex="male and female"
                       /dev_stage="adult"
                       /lab_host="DH5 - alpha"
                       /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
                       XhoI; Sized fractionated cDNAs were directly ligated into
                       pOT2. Plasmid cDNA library."         1 others
BASE COUNT             166 a       138 c       179 g       146 t
ORIGIN

Query Match            2.9%;      Score 56.2;  DB 9;   Length 630;
Best Local Similarity  53.48%;    Pred. No. 0.0005;
Matches 118;   Conservative    0;  Mismatches 103;  Indels  0;  Gaps   0;

Qy     234  caacggtgagatctacaactacgttgagctgcgtaaagagctctcggattgggatatac 293
            |||| |||| ||||| |||||||  ||||||||||| ||| ||  |||||||| |||||
Db     330  CAATGGAGAGATCTACAATTCTCGGAGCTTTCGGCTGAGATTCCGAAAAAGCGTGGCTC 389

Qy     294  ctttaatacttctggcgatggcgagccaatgttgtcggtttccaccactgggcgagtc 353
            ||| | |||  ||||||||| | |||||  ||| ||| | ||| | ||||||| |||
Db     390  CTACAACCCCATGAGCGATTGCCACGTGATACTGGAACTGTACCACGGATTATGGAAAGGA 449

Qy     354  cgtggtcgagcatctcccgcggaatgttcggcattgccattgggatacaaaggaaaagtc 413
            | ||||||||||  |||| |  ||| ||||| |||||  ||||  | |||| ||| |||
Db     450  TCTTCTGCAATACATCACTGGAATGTTTGCTTTGCCTTGCTGTCGCGATAGGAAGACCAAGGA 509

Qy     414  gctttcctgcgcgtgatcagttcggcatcaagccactgt 454
            |||  |||| | ||| |||| ||| ||| | ||| |||
Db     510  AGTGCTTCTTGCCCGAGACCCCTTTGGCATTATACCCATGT 550
```

```
RESULT 11
LOCUS       BI635460             644 bp    mRNA    linear    EST 10-SEP-2001
DEFINITION  SD16825.5prime SD Drosophila melanogaster Schneider L2 cell culture
            pOT2 Drosophila melanogaster cDNA clone SD16825 5, mRNA sequence.
ACCESSION   BI635460
VERSION     BI635460.1  GI:15537670
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 644)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
```

```
                       Fax: 510 486 6798
                       Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
                       hit genomic AE003593: arm:3L [20799555,21112669]
                       estimated-cyto:78A2-78C7; 05/18/2001
                       Plate: SD.168  row: C  column: 1
                       High quality sequence stop: 548.
FEATURES               Location/Qualifiers
  source               1..644
                       /organism="Drosophila melanogaster"
                       /db_xref="taxon:7227"
                       /clone="SD16825"
                       /clone_lib="SD Drosophila melanogaster Schneider L2 cell
                       culture pOT2"
                       /lab_host="DH5-alpha"
                       /note="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
                       fractionated cDNAs were directly ligated into pOT2.
                       Plasmid cDNA library."
BASE COUNT             166 a       143 c       177 g       158 t
ORIGIN

Query Match            2.9%;       Score 56.2;  DB 10;   Length 644;
Best Local Similarity  53.48%;     Pred. No. 0.00051;
Matches 118;   Conservative    0;  Mismatches 103;   Indels   0;  Gaps   0;

Qy     234  caacggtgagatctacaactacgttgagctgcgtaaagagctctcggattgggatatac 293
            |||| |||| ||||| |||||||  ||||||||||| ||| ||  |||||||| |||||
Db     314  CAATGGAGAGATCTACAATTCTCGGAGCTTTCGGCTGAGATTCCGAAAAAGCGTGGCTC 373

Qy     294  ctttaatacttctggcgatggcgagccaatgttgtcggtttccaccactgggcgagtc 353
            ||| | |||  ||||||||| | |||||  ||| ||| | ||| | ||||||| |||
Db     374  CTACAACCCCATGAGCGATTGCCACGTGATACTGGAACTGTACCACGGATTATGGAAAGGA 433

Qy     354  cgtggtcgagcatctcccgcggaatgttcggcattgccattgggatacaaaggaaaagtc 413
            | ||||||||||  |||| |  ||| ||||| |||||  ||||  | |||| ||| |||
Db     434  TCTTCTGCAATACATCACTGGAATGTTTGCTTTGCCTTGCTGTCGCGATAGGAAGACCAAGGA 493

Qy     414  gctttcctgcgcgtgatcagttcggcatcaagccactgt 454
            |||  |||| | ||| |||| ||| ||| | ||| |||
Db     494  AGTGCTTCTTGCCCGAGACCCCTTTGGCATTATACCCATGT 534
```

```
RESULT 12
BI142084
LOCUS       BI142084             693 bp    mRNA    linear    EST 03-JUL-2001
DEFINITION  SD16082.5prime SD Drosophila melanogaster Schneider L2 cell culture
            pOT2 Drosophila melanogaster cDNA clone SD16082 5, mRNA sequence.
ACCESSION   BI142084
VERSION     BI142084.1  GI:14594528
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 693)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003593: arm:3L [20799555,21112669]
            estimated-cyto:78A2-78C7: 04/13/2001
            Plate: SD.160  row: G  column: 10
            High quality sequence stop: 604.
FEATURES               Location/Qualifiers
  source               1..693
                       /organism="Drosophila melanogaster"
```

ORIGIN

```
Query Match          2.9%;   Score 56.2;   DB 10;   Length 701;
Best Local Similarity  53.4%;  Pred. No. 0.00053;
Matches  118;   Conservative    0;   Mismatches 103;   Indels    0;   Gaps    0;

Qy    234  caacggtgagatctcaaactacgttgagctgcgtaaagagctctcggattgggatatac 293
           ||||||||||||||||||||| || |||||||  | |||||| || ||||||| |||||
Db    321  CAATGGAGAGAGATCTACAATTATCTGGAGCTTTCGGCTGAGATTGCGAAAAAGCGTGGCTC 380

Qy    294  ctttaatacttctggcgatggcgagccaattgttgtcggttccaccactgggcggagtc 353
           || ||| |||||||||||||||||||||  ||| ||  |||||| || |||||| ||||
Db    381  CTACAACCCCATGAGCGATTGCCACGTGATACTGGAACTGTACCAGGATTATGGAAAGGA 440

Qy    354  cgtggtcgagcatccgcggaatgttcggcattgccattgggatacaaaggaaaagtc 413
           ||||||||||||||||||||||| ||||||||||||| ||||| ||| ||| ||||
Db    441  TCTTCTGCAATACATCACTGGAATGTTTGCTTTGCCTTGTGTACGATAGGAAGACCAAGGA 500

Qy    414  gcttttccttgcgcgtgatcagttcggcatcaagccactgt 454
           |||||||||||||||||||||| |||||| |||||||||
Db    501  AGTGCTTCTTGCCCGAGACCCCTTTGGCATTATACCCATGT 541
```

RESULT 14
BI719579
LOCUS   BI719579   640 bp   mRNA   linear   EST 19-SEP-2001
DEFINITION   1031044E03.y1 C. reinhardtii CC-1690, Stress II (normalized),
             Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BI719579
VERSION   BI719579.1   GI:15695274
KEYWORDS   EST.
SOURCE   Chlamydomonas reinhardtii
  ORGANISM   Chlamydomonas reinhardtii
             Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
             Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1 (bases 1 to 640)
  AUTHORS   Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
             ,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
  TITLE   Analyses of the Chlamydomonas reinhardtii Genome: A Model,
             Unicellular System for Analyzing Gene Function and Regulation in
             Vascular Plants. Project: 1031
  JOURNAL   Unpublished (2001)
  COMMENT   Contact: Charles Hauser
             DCMB Box 91000
             Duke University
             Durham, NC 27708-1000
             Tel: 919 613 8159
             Fax: 919 613 8177
             Email: chauser@duke.edu.
FEATURES   Location/Qualifiers
     source   1..640
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, Stress II (normalized
                     ), Lambda Zap II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI: Stress condition II library, constructed by John
                     Davies and Jeffrey McDermott, combines cDNAs from CC-1690
                     cells grown to mid-log phase in TAP (NH4+ - containing)
                     and shifted to TAP - NO3- (24hrs); H2 production
                     conditions (0, 12hr, 24hr) see Melis et al.,(2000) plant
                     Phys. 122; 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
                     sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized.The cDNA was directionally cloned into lambda
                     Zap II (Stratagene) in the EcoRI (5') and XhoRI (3')
                     sites. pBluescript II SK- plasmids were excised from the
                     lambda ZAP clones by superinfection with ExAssist
                     (Stratagene) phage. The library was normalized using
                     method 4 described in Bonaldo et al., (1996) Genome
                     Research 6: 791-806."

BASE COUNT   177 a   150 c   194 g   172 t
ORIGIN
```

```
/db_xref="taxon:7227"
/clone="SD16082"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell
culture pOT2"
/lab_host="DH5-alpha"
/note="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
fractionated cDNAs were directly ligated into pOT2.
Plasmid cDNA library."
```

BASE COUNT   177 a   150 c   194 g   172 t
ORIGIN

```
Query Match          2.9%;   Score 56.2;   DB 10;   Length 693;
Best Local Similarity  53.4%;  Pred. No. 0.00053;
Matches  118;   Conservative    0;   Mismatches 103;   Indels    0;   Gaps    0;

Qy    234  caacggtgagatctacaactacgttgagctgcgtaaagagctctcggattgggatatac 293
           ||||||||||||||||||||| || |||||||  | |||||| || ||||||| |||||
Db    316  CAATGGAGAGAGATCTACAATTATCTGGAGCTTTCGGCTGAGATTGCGAAAAAGCGTGGCTC 375

Qy    294  ctttaatacttctggcgatggcgagccaattgttgtcggttccaccactgggcggagtc 353
           || ||| |||||||||||||||||||||  ||| ||  |||||| || |||||| ||||
Db    376  CTACAACCCCATGAGCGATTGCCACGTGATACTGGAACTGTACCAGGATTATGGAAAGGA 435

Qy    354  cgtggtcgagcatccgcggaatgttcggcattgccattgggatacaaaggaaaagtc 413
           ||||||||||||||||||||||| ||||||||||||| ||||| ||| ||| ||||
Db    436  TCTTCTGCAATACATCACTGGAATGTTTGCTTTGCCTTGTGTACGATAGGAAGACCAAGGA 495

Qy    414  gcttttccttgcgcgtgatcagttcggcatcaagccactgt 454
           |||||||||||||||||||||| |||||| |||||||||
Db    496  AGTGCTTCTTGCCCGAGACCCCTTTGGCATTATACCCATGT 536
```

RESULT 13
BI641950
LOCUS   BI641950   701 bp   mRNA   linear   EST 10-SEP-2001
DEFINITION   SD25636.5prime SD Drosophila melanogaster Schneider L2 cell culture
             pOT2 Drosophila melanogaster cDNA clone SD25636 5, mRNA sequence.
ACCESSION   BI641950
VERSION   BI641950.1   GI:15544160
KEYWORDS   EST.
SOURCE   fruit fly.
  ORGANISM   Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
             Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
             Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 701)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
             Lewis,S. and Rubin,G.M.
  TITLE   BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
  COMMENT   Contact: Stapleton, M.
             BDGP
             Lawrence Berkeley National Lab
             One Cyclotron Rd, Berkeley, CA 94720, USA
             Fax: 510 486 6798
             Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
             hit genomic AE003593: arm:3L [20799555,21112669]
             estimated-cyto:78A2-78C7: 05/19/2001
             Plate: SD.256 row: C column: 12
             High quality sequence stop: 574.
FEATURES   Location/Qualifiers
     source   1..701
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="SD25636"
                     /clone_lib="SD Drosophila melanogaster Schneider L2 cell
                     culture pOT2"
                     /lab_host="DH5-alpha"
                     /note="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
                     fractionated cDNAs were directly ligated into pOT2.
                     Plasmid cDNA library."

BASE COUNT   182 a   151 c   195 g   172 t   1 others
```

```
BASE COUNT    129 a    201 c    185 g    125 t
ORIGIN

Query Match    2.9%;    Score 55;    DB 10;    Length 640;
Best Local Similarity  47.4%;    Pred. No. 0.0011;
Matches  246;    Conservative  0;    Mismatches  255;    Indels  18;    Gaps  2;

Qy      1 atgtgcggccttcttggcatattgactgcaaatgggaaacgctgaagcattcgttcctgca  60
          |||||||||| |||||  |||| ||||| ||  ||||  ||||||||||||| |||| |||
Db     87 ATGTGCGGCATTCTTGCCGTCCTCAACACACGATGACAGCCAGGCTATGCCGCTCGAGG 146

Qy     61 ctcgagcgggccttgccatgcatgcgccaccgtggtcctgacgatgccggcacttggcat 120
          ||                ||| ||||||||||||||||| |||
Db    147 GTGCTGGCCCTGAGCCGTCGCCAGCGTCACCGTGCGCCCGACTGGTCTCGGCATGCACCAG 206

Qy    121 gacgccgatgcagcgtttgattcaaccgcctcccatcattgatattgcacactcccac 180
          ||            |||| ||| |||| ||| |||||||||||||| ||||
Db    207 TTCGGCAACAACTTCCTTGCCATGAGCGCCTTGGCATTATGGACCCCGCCTCGGGTGAC 266

Qy    181 caaccatgcgttggggacctgcggatgaaccgccgccgcactacgcaatgactttcaacggt 240
          |||| |||                     ||  ||||||||||||||||||||||||
Db    267 CAGCCCCTG---------TTCAACGAGGACCGCCACAATCGTGGTCACCGTGAACGGT 314

Qy    241 gagatctacaactacgttgagctcgtaaagagctctcggatttg------ggatatacc 294
          |||||||||||||  |||||||| ||||||||||||
Db    315 GAGATCTACAACTACAAGGAGCTGCGCCAGCGAGATGCCTGCCCCGGCAAGAAG 374

Qy    295 tttaatacttctgggcgatggcgagccaattgttgtcggttccaccactggggcgagtcc 354
          ||||| ||||||||||||     |||||||||||||||||||||||||||
Db    375 TTCGGCCACCAACAGCGATTGCGAGGTGATTAGCCACCTGTACGAGCTGACGCGAGAAG 434

Qy    355 gtggtcgagcatctccggaatgttcggcattgggatacaaaggaaaagtcg 414
          |||||||||||||| |||||||||||||||||||||||||||||||||
Db    435 GTGGGCCTCCATGCTGGACGGCTTCTTCGCTTCGTGGTGCTGGACCACCCCGAACAACACC 494

Qy    415 ctttccttgcgtgatcagttcggcatcaagccactgttctacgcaaccaccgagcat 474
          ||||       || ||||| || |||||||||||||
Db    495 TTCTACGCCGGCGGCCGGCAGCCCCATTGGCATGCATCGATGTACATCGGCTGGGGCCGTGAC 554

Qy    475 ggcaccgtgttctcctcagagaagaagaccatcttggag 513
          ||||| || ||  || || |||
Db    555 GGCAGCGTGTGGCGTCGGCGTGTCGAGCGAGATGAAGTGCCTGAAG 593
```

```
RESULT 15
LOCUS        BI354992               620 bp    mRNA       linear   EST 31-JUL-2001
DEFINITION   GM29151.5prime GM Drosophila melanogaster ovary pOT2 Drosophila
             melanogaster cDNA clone GM29151 5, mRNA sequence.
ACCESSION    BI354992
VERSION      BI354992.1  GI:15049446
KEYWORDS     EST.
SOURCE       fruit fly.
ORGANISM     Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
             Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
             Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1  (bases 1 to 620)
AUTHORS      Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
             Lewis,S. and Rubin,G.M.
TITLE        BDGP/HHMI Drosophila EST Project
JOURNAL      Unpublished (2001)
COMMENT      Contact: Stapleton, M.
             BDGP
             Lawrence Berkeley National Lab
             One Cyclotron Rd, Berkeley, CA 94720, USA
             Fax: 510 486 6798
             Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
             hit genomic AE003593::arm:3L [20799555,21112669]
             estimated-cyto:78A2-78C7: 05/22/2001
             Plate: GM.291 row: E  column: 3
             High quality sequence stop: 579.
```

```
FEATURES             Location/Qualifiers
     source          1..620
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="GM29151"
                     /clone_lib="GM Drosophila melanogaster ovary pOT2"
                     /sex="female"
                     /dev_stage="newly eclosed females: germarium-stage 6"
                     /lab_host="XL1 Blue"
                     /note="Organ: ovary; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2."

BASE COUNT    164 a    134 c    174 g    148 t
ORIGIN

Query Match    2.8%;    Score 54.6;    DB 10;    Length 620;
Best Local Similarity  52.9%;    Pred. No. 0.0013;
Matches  117;    Conservative  0;    Mismatches  104;    Indels  0;    Gaps  0;

Qy    234 caacggtgagatctacaactacgttgagctcgtaaagagctctcggatttgggatatac 293
          ||||||||||||||||||||||||  ||  || ||||||||||
Db    320 CAATGGAGAGATCTACAATTATCTGGAGCTTTCGGCTGAGATTGCGAAAAAGCGTGGCTC 379

Qy    294 ctttaatacttctggcgatggcgagccaattgttgtcggttccaccactgggcggcgagtc 353
          ||  ||||||||| ||  ||||||||||||||||||||||||
Db    380 CTACAACCCCATGAGGCGATTGCCACGTGATACTGACCTACCAGGATTATCGAAAGGA 439

Qy    354 cgtggtcgagcatctccggaatgttcggcattgtcggcattggggatacaaaggaaaaagtc 413
          || |||||||||||| |||||||||||||||||||||||||
Db    440 TCTTCTGCAATACATCACTGGAATGTTTGCTTTGCCTTGTACGATAGGAAAACCAAGGA 499

Qy    414 gctttccttgcgtgatcagttcggcgtcgatcaagccactgt 454
          |||||||||||||||||||| || ||||||||||||
Db    500 AGTGCTTCTTGCCCGAGCACCCCTTGGCATTATACCCATGT 540
```

Search completed: September 20, 2002, 06:29:38
Job time: 6070 sec

OM of: US-09-786-474-2 to: N_Geneseq_032802:*   out_format : pfs
Date: Sep 20, 2002   9:35 PM
About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+p2n.model -DEV=xlh
 -O=/cgn2.1/USPTO_spool/US09786474/runat_18092002_132913_5515/app_query.fasta_1.706
 -DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=p2n.rng -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=10 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09786474 -@CGN1_1_0 -NCPU=6 -ICPU-3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-786-474-2
Query length: 640
Database: N_Geneseq_032802:*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 212.170000

score_list:

Sequence	Strd	Orig	ZScore	EScore	EScore	Len	! Documentation
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA10292	+	3379.00	5746.06	1.0e-311		19	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA10293	+	3379.00	5738.68	2.7e-311		38	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH67371	+	3366.00	5723.85	1.8e-310		1	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH68531	+	3366.00	5668.07	2.3e-307		3	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:AAF71807	+	3360.00	5713.37	6.9e-310		4	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:AAT61016	-	633.00	1025.18	9.4e-49		30	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAX05110	-	633.00	1025.18	9.4e-49		30	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF77894	+	605.00	1007.07	9.5e-48		1	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH41225	-	560.50	875.00	2.2e-40		3	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA58471	+	516.00	818.07	3.2e-37		58	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC40911	+	468.00	771.46	1.3e-34		21	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC50959	+	467.00	769.76	1.6e-34		21	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAV81946	-	463.50	732.48	1.9e-32		40	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAA71520	-	463.50	732.48	1.9e-32		40	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAI64984	-	463.50	732.10	2.0e-32		4	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC46946	+	452.50	746.74	3.0e-33		18	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV21209	+	439.00	650.71	6.7e-28		16	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:AAQ06598	+	428.00	702.95	8.3e-31		22	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:ABI99327	+	422.50	694.96	2.3e-30		19	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAC78066	+	418.50	687.57	6.0e-30		20	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH34419	+	418.00	687.57	6.0e-30		3	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH41225	+	417.00	629.83	9.8e-27		3	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS68243	-	379.00	607.98	1.6e-25		6	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF60005	+	192.00	299.01	2.6e-08		2	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAI58275	+	192.00	297.54	3.2e-08		2	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAI60061	+	190.00	295.36	4.2e-08		2	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:AAT09700	+	184.00	271.80	8.6e-07		83	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS54024	+	172.50	268.41	1.3e-06		1	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS56330	+	171.00	265.84	1.8e-06		1	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS41217	+	166.00	268.26	1.4e-06		6	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL19004	-	166.00	249.74	1.5e-05		3	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAZ60717	+	165.00	257.43	5.4e-06		15	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAD14504	+	165.00	257.43	5.4e-06		1	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAD14501	+	165.00	235.80	8.7e-05		1	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:AAQ91580	+	165.00	233.00	0.0001		15	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAD14499	+	164.50	233.00	0.0001		1	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS52665	+	164.50	254.74	7.7e-06		1	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAZ58252	+	164.50	252.84	9.8e-06		21	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAZ58253	+	164.50	252.84	9.8e-06		21	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAZ58254	+	164.50	252.84	9.8e-06		21	

/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAZ58249 + 164.50 239.76 5.2e-05
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL19005 + 163.00 252.94 9.7e-06
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAZ58250 + 162.50 249.43 1.5e-05
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAZ58251 + 160.50 246.01 2.4e-05
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ01425 - 160.50 179.96 0.1122

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA10292

seq_documentation_block:
ID   AAA10292 standard; DNA; 1920 BP.
XX
AC   AAA10292;
XX
DT   03-JUL-2000 (first entry)
XX
DE   DNA encoding Corynebacterium glutamicum lysozyme insensitivity protein.
XX
KW   Lysozyme insensitivity protein; bacterium; recombinant expression;
KW   amino acid production; ds.
XX
OS   Corynebacterium glutamicum.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..1920
FT   /*tag=   a
FT   /product= "Corynebacterium glutamicum lysozyme
FT   insensitivity protein"
FT   /note= "No stop codon given in the specification"
XX
PN   WO20001424l-Al
XX
PD   16-MAR-2000
XX
PF   04-SEP-1998;   98WO-JP03981.
XX
PR   04-SEP-1998;   98WO-JP03981.
XX
PA   (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI   Nagai K, Wat1 M;
XX
DR   WPI; 2000-256989/22.
DR   P-PSDB; AAY87459.
XX
PT   New DNA molecule useful for production of amino acids encodes protein
PT   imparting lysozyme resistance to Corynebacterium glutamicum strain -
XX
PS   Claim 3; Page 33-35; 50pp; Japanese.
XX
CC   This sequence represents the DNA coding sequence encoding a
CC   Corynebacterium glutamicum lysozyme insensitivity protein. This protein
CC   renders lysozyme-sensitive strains of Corynebacterium glutamicum
CC   insensitive to lysozyme. DNA sequences encoding the lysozyme
CC   insensitivity protein can be used to construct a recombinant vector for
CC   the expression of this protein in a host cell. The DNA encoding the
CC   lysozyme insensitivity protein can also be mutated, and used to generate
CC   strains of Corynebacterium glutamicum in which this protein is
CC   inactivated and which are therefore lysozyme-sensitive. Corynebacterium
CC   glutamicum is used for the production of amino acids, in particular
CC   glutamic acid and glutamine.
XX
SQ   Sequence 1920 BP; 425 A; 535 C; 524 G; 436 T; 0 other;

alignment_scores:
         Quality: 3379.00   Length: 640
         Ratio: 5.280   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-786-474-2 x AAA10292

Align seg 1/1 to: AAA10292   from: 1   to: 1920

```
  1 MetCysLysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaAlaPh   17
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  1 ATGTGCAAGGGCCTTCTTGGCATATTGACTGCAAATGGGAACGCTGAAGCATT     50

 17 eValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA        34
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 CGTTCCTGCACTCGAGCGGGCCTTGCCATGCGACGCCACCGTGGTCCTG       100

 34 spAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGlyPheAsnArg        50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 ACGATGCGCCGACCACTTGGACATGACGACGCCGATGCGTTTGGAATCAACCGC  150

 51 LeuSerIleIleAspHisSerSerHisGlnProLeuProArgLeuTrpGlyPr     67
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 CTCTCCATCATTGATATTGACACTCCCACCAACCACTGCGTTGGGACC       200

 67 oAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyIleTyrA          84
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 TGCGGATGAACCCGACCGCTACGCAGTGACCTTCAACGGTGAGATCTACA      250

 84 snTyrValGlyLeuArgLysGluGluAspAspLeuGlyTyrThrPheAsn      100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
251 ACTACGGTGAGCGTGCTAAAGAGCTCTCGGATTTGGGATATACCTTTAAT     300

100 ThrSerGlyAspGlyAspProIleValValGlyPheHisHisTrpGlyG1      117
    ||||||||||||||||||||||||||||||||||||||||||||||||||
301 ACTTCTGGCGATGGCGAGCCAATTGTTGTCGGTTTCCACCACTGGGGCGA     350

117 uSerValValGluHisIleLeuArgGlyMetPheGlyIleLeuAlaIleTrpAspT 134
    ||||||||||||||||||||||||||||||||||||||||||||||||||
351 GTCCGTGGTCGAGCATCTTCCGCGCGGCATGTTCGGCATTGCCATTGGGATA   400

134 hrLysGluLysSerLeuPheLeuAlaArgAspGlnPheGlyIleLeuLysPro   150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
401 CAAGGAAAAGTCGCTTTTCCTTGCGCGTGATCAGTTCGGCATCAAGCCA     450

150 LeuPheTyrAlaThrThrGluHisGlyTyrThrValPheSerSerGluLysLysLy 167
    ||||||||||||||||||||||||||||||||||||||||||||||||||
451 CTGTTCTACGCAACCACCGAGCACGGCTACACCGTGTTCTCCTCAGAGAAGAA  500

167 sThrIleLeuGluMetAlaGluLeuMetAsnLeuAspLeuGlyLeuAspL      184
    ||||||||||||||||||||||||||||||||||||||||||||||||||
501 GACCATCTTGGAGATGGCCGAGCTGATGAATCTAGATCTGGGCCTTGATA     550

184 ysArgThrIleGlyHisTyrValAspLeuGluGlnTyrValProGluProAsp   200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
551 AGCGCACCATTGAGCACTACGTGGACCTGCAGTACGTGCCCGAGCCAGAT     600

200 ThrLeuHisAlaGlnIleIleSerArgLeuSerGluSerGlyThrAlaThrVa   217
    ||||||||||||||||||||||||||||||||||||||||||||||||||
601 ACCCTGCACGCGCAGATTCCGAGTCGAGGCTGCAGGCTGCACCGCAACAGT   650

217 lArgProGlyLysLeuGluGlnLysLysArgTyrPheLysProGlnPheP      234
    ||||||||||||||||||||||||||||||||||||||||||||||||||
651 TCGTCCGGGCGGCAAGCTGGAACAGAAGAAGCGTTACTTCAAGCCTCAGTTCC 700

234 roValGlnLysLysGlyValValLysGlyLysGluGlnAspPheAspArgIle  250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
701 CAGTACAGAAGAAGGGTCGTAAAGGGTAAGGAGCAGGACTTCGATCGCATT   750

250 AlaGlnValLeuGluAsnProSerValGluLysLysMetArgAlaAlaAspValTh 267
    ||||||||||||||||||||||||||||||||||||||||||||||||||
751 GCCCAGGTGTTGGAGAATAGCGTCGAAAAGCATATGCGTGCCGACGTGAC    800

267 rValGlySerPheLeuSerGlyLeuIleLeuAspProSerThrAlaIleAlaProL 284
    ||||||||||||||||||||||||||||||||||||||||||||||||||
801 CGTAGGCTCGTTTCTTTCCGGCCTCATTCAACCGCAATTGCGCCCGGC    850

284 euAlaLysArgHisAsnProAspLeuLysLeuLysLeuThrProPheThrThrGlyGlu 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
851 TTGCAAAGCGCCAACCCTGACCTGCTCACCCTTCACCACCGGTTTCGAG    900
```

```
301 ArgGluGlyTyrSerGluValAspValAlaAlaAlaGluSerAlaAlaAlaIl     317
    ||||||||||||||||||||||||||||||||||||||||||||||||||
901 CGTGAAGGCTACTCGGAGGTCGATGTGGCTGCGGCGGAGTCCGCCGCTGCGAT   950

317 eGlyAlaGluHisIleValLysIleValLysIleValSerProGluGlyTyrAlaAsnA 334
    ||||||||||||||||||||||||||||||||||||||||||||||||||
951 TGGCGCTGAGCACATCGTGAAGATTGTCTCGCCTGAGGAATACCGCCAACG   1000

334 IaIleProLysIleMetTrpTyrLeuAspAspProValAlaAspProSer     350
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1001 CGATTCCTAAGATCATGTGGTACTTGGATGATCCTGTAGCTGACCCATCA   1050

351 LeuValProLeuTyrPheValAlaAlaGluAlaArgLysHisValLysVa    367
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1051 TTGGTCCCCGTGTACTTCGTGGCAGCGGAGGCACGTAAGCACGTCAAGGT  1100

367 lValLeuSerGlyGluGlyAlaAspGluLeuPheGlyGlyTyrThrIleT   384
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1101 TGTGCTGTCTGGCGAGGGCGCAGATGAGCTTCGGTGGATACACCATTT   1150

384 yrLysGluProLeuSerLeuAlaProPheGluLysIleProSerProLeu  400
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1151 ACAAGAGCCGCTATCGCTTGCTCCATTTGAGAGATCCCTTCCCCACTA   1200

400 ArgLysGlyLeuGlyLysLeuLysSerLysValLeuProAspGlyMetLysGl 417
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1201 CGTAAAGGCCTGGGAAAGCTCAGCAAGGTTCTGCCAGACGGCATGAAGGG  1250

417 yLysSerLeuLeuGluArgGlySerMetThrMetGluGluArgTyrTyrG   434
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1251 CAAGTCCCTTCTTGAGCGTGGCTCCATGACCATGGAAGAGCGCTACTACG  1300

434 lyAsnAlaArgSerPheAsnPheGluGlnMetGlnArgValIleProTrp   450
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1301 GCAACGCCTCGCTCCTTCAATTCGAGCAGATGCAACGCGTTATTCCATGG  1350

450 AlaLysArgGluTrpAspHisArgGluValThrAlaProIleTyrAlaGl   467
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1351 GCAAAGCGCGAATGGACCACCGCGAAGTCACTGCACCGATCTACGCACA  1400

467 nSerArgAsnPheAspPheAsnPheValAlaArgMetGlnHisLeuAspLeuPheT 484
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1401 ATCCCGCAACTTTGATCCAGTAGCCCGCATGCAACACCTGGATCTGTTCA  1450

484 hrTrpMetArgGlyAspIleLeuValLysAlaAspLysIleAsnMetAla   500
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1451 CCTGGATGCGCGGCGACATCCTGGTCAAGGCTGACAAGATCAACATGGCG  1500

501 AsnSerLeuLeuArgValProPheLeuAspLysGluValPheLysVa     517
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1501 AACTCCCTTGAGCTGCGAGTTCCATTCTTGGATAAGGAAGTTTTCAAGGT  1550

517 lAlaGluThrIleProTyrAspLeuLysIleAlaAsnGlyThrThrLysT   534
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1551 TGCAGAGACCATTCCTTACGATCTGAAGATTGCCAACGGTACCACCAAGT  1600

534 yrAlaLeuArgArgAlaLeuGluGluIleValProProHisValLeuHis  550
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1601 ACGCGCTGCGGCAGGCACTCGAGCAGATTGTTCCGCCTCACGTTTGCAC   1650

550 ArgLysLysLeuPheProValProMetArgHisTrpLeuAlaGlyAs     567
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1651 CGCAAGAAGCTGGCTTCCCATGCGCCACTGGCTTGCCGGCGGA       1700

567 pGluLeuPheGlyTrpAlaGlnAspThrIleLysGlySerGlyThrGluA   584
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1701 TGAGCTGTTCGGTTGGGCGCAGGACACCATTAAGGAATCCGGTACTGAAG  1750

584 spIlePheAsnLysGlnAlaValLeuAspMetLeuAsnGluHisArgAsp  600
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1751 ATATCTTCAACAAGCAGGCTGTGCTGGATATGCTGAACGAGCACCGCGAT  1800

601 GlyValSerAspHisSerArgArgLeuTrpThrValLeuSerPheMetVa   617
```

```
1801 GGCGTGTCAGATCATTCCCGTCGACTGTGGACTGTTCTGTCATTTATGGT 1850
       ||||||||||||||||||||||||||||||||||||||||||||||||||
 617 lTrpHisGlyIlePheValGluAsnArgIleAspProGlnIleGluAspA 634
 1851 GTGGCACGGCCATTTTGTGGAAAACCGCATTGATCCACAGATTGAGGACC 1900

 634 rgSerTyrProValGluLeu 640
       ||||||||||||||||||||
 1901 GCTCCTACCCGGTCGAGCTT 1920
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA10293

seq_documentation_block:
ID   AAA10293 standard; DNA; 3825 BP.
XX
AC   AAA10293;
XX
DT   03-JUL-2000 (first entry)
XX
DE   DNA encoding Corynebacterium glutamicum lysozyme insensitivity protein.
XX
KW   Lysozyme insensitivity protein; bacterium; recombinant expression;
KW   amino acid production; ds.
XX
OS   Corynebacterium glutamicum.
XX
FH   Key   Location/Qualifiers
FT   CDS   815..2737
FT   /tag= a
FT   /product= "Corynebacterium glutamicum lysozyme
FT   insensitivity protein"
XX
PN   WO20001424l-Al.
XX
PD   16-MAR-2000.
XX
PF   04-SEP-1998; 98WO-JP03981.
XX
PR   04-SEP-1998; 98WO-JP03981.
XX
PA   (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI   Nagai K, Wati M;
XX
DR   WPI; 2000-256989/22.
DR   P-PSDB; AAY87459.
XX
PT   New DNA molecule useful for production of amino acids encodes protein
PT   imparting lysozyme resistance to Corynebacterium glutamicum strain -
XX
PS   Example 2; Page 40-44; 50pp; Japanese.
XX
CC   This sequence represents DNA encoding a Corynebacterium glutamicum
CC   lysozyme insensitivity protein. This protein renders lysozyme-sensitive
CC   strains of Corynebacterium glutamicum insensitive to lysozyme. DNA
CC   sequences encoding the lysozyme insensitivity protein can be used to
CC   construct a recombinant vector for the expression of this protein in a
CC   host cell. The DNA encoding the lysozyme insensitivity protein can also
CC   be mutated, and used to generate strains of Corynebacterium glutamicum in
CC   which this protein is inactivated and which are therefore lysozyme-
CC   sensitive. Corynebacterium glutamicum is used for the production of amino
CC   acids, in particular glutamic acid and glutamine.
XX
SQ   Sequence 3825 BP; 881 A; 1031 C; 993 G; 920 T; 0 other;

alignment_scores:
          Quality: 3379.00   Length: 640
            Ratio: 5.280   Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-786-474-2 x AAA10293

Align seg 1/1 to: AAA10293 from: 1 to: 3825

```
   1 MetCysGlyLeuLeuGlyIleLeuThrLeuAsnGlyAsnAlaGluAlaPh  17
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 815 ATGTGCGGCCTTCTTGGCATATTGACTCTGAACGGGAACGCTGAAGCATT  864

  17 eValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA  34
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 865 CGTTCCTGCACTCGAGCGGGCCTTGCCATGCATGCGCCACCGTGGTCCTG  914

  34 spAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGlyPheAsnArg  50
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 915 ACGATGCCGGCACTTGGCATGACGCGGCCGATGACGCGTTTGGATTCAACCGC  964

  51 LeuSerIleIleAspIleAlaHisSerHisGlnProLeuArgTrpGlyPr  67
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 965 CTCTCCATCATTGATATTGCACACTCCCACCACTCGCGTTGGGGACC  1014

  67 oAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyGlyGluTyrA  84
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1015 TGCGGATGAACCGCCATCCTACGGCATGACTTTCAACGGTGAGATCTACA  1064

  84 snTyrValGlyLeuGluArgLysGluLeuSerAspLeuGlyTyrThrPheAsn 100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1065 ACTACGTTGAGCTGCGTAAAGAGCTCTCGGATTTGGGATATACCTTTAAT 1114

 101 ThrSerGlyAspGlyGlyProIleValValGlyPheHisHisTrpGlyG1 117
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1115 ACTTCTGGCGATGGCGAGCCAATTGTTGTCGGTTTCCACCACTGGGGCGA 1164

 117 uSerValGluHisLeuArgGlyMetPheGlyIleAlaIleTrpAspT 134
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1165 GTCCGTGGTCGAGCATCTCCGCGGAATGTTCGGCATTGCCATTTGGGATA 1214

 134 hrLysGluLysSerLeuPheLeuAlaArgAspAspIleLysPheLysPro 150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1215 CAAAGGAAAAGTCGCTTTTCCTTGCGCGTGATCAGTTCGGCATCAAGCCA 1264

 151 LeuPheTyrAlaThrThrGluHisGlyThrValPheSerSerGluLysLy 167
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1265 CTGTTCTACGCAACCACCGAGCATGGCACCGTGTTCTCCTCAGAGAAGAA 1314

 167 sThrIleLeuGluMetAlaGluGluMetAsnLeuAsnLeuGlyLeuAspL 184
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1315 GACCATCTTGGAGATGGCCGAGGAGATGAATCTAGATCTGGGCCTTGATA 1364

 184 ysArgThrIleGluHisTyrValAspLeuGlnTyrValProGluProAsp 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1365 AGCGCACCATTGAGCACTACGTGGACCTGCAGTACGTGCCCGAGCCAGAT 1414

 201 ThrLeuHisAlaGlnIleSerArgLeuGluSerGlyCysThrAlaThrVa 217
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1415 ACCCTTCACGCGCAGATTTCCCGCTTGGAGTCAGGCTGCACCGCAACAGT 1464

 217 lArgProGlyLysLeuGluGlnLysArgTyrPheLysProGlnPheP 234
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1465 TCGTCCGGGCGGCAAGCTGGAACAGAAGCGTTACTTCAAGCCTCAGTTCC 1514

 234 roValGlnLysValValLysGlyLysGluGlnAspLeuPheAspArgIle 250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1515 CAGTACAGAAGGTCGTAAAGGGTAAGGACGCAGGACCTCTTCGATCGCATT 1564

 251 AlaGlnValLeuGluAspSerValGluLysHisMetArgAlaAspValTh 267
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1565 GCCCAGGTGTTGGAGGATAGCGTCGAAAAGCATATGCGTGCCGACGTGAC 1614

 267 rValGlySerPheLeuSerGlyIleIleAspSerThrAlaIleAlaProL 284
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1615 CGTAGGCTCGTTCCTTTCCGGCGGCATTGACTCAACCGCAATTGCGCCGC 1664

 284 euAlaLysArgHisAsnProAspLeuThrPheThrPheThrGlyPheGlu 300
```

```
1665 TTGCAAAGCGCCACAAACCCTGACCCTGCTCACCTTCACCACCGGTTTCGAG 1714
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 301 ArgGluGlyTyrSerGluValAspValAlaAlaAlaGluSerAlaAlaAlaIl 317
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1715 CGTGAAGGCTACTCGGAGGTCGATGTGGCTGCCGCCGAGTCCGCCGCTGCCAT 1764
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 317 eGlyAlaGluHisLysIleValLysIleValSerProGluGluTyrAlaAsnA 334
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1765 TGGCGCTGAGCACATCGTGAAGATTGTCTCGCCTGAGGAATACGCCAACG 1814
     |||||||||||||||||||||||||||||||||||||||||||||||||
 334 lAIleProLysIleMetTrpTyrLeuAspAspProValAlaAspProSer 350
     |||||||||||||||||||||||||||||||||||||||||||||||||
1815 CGATTCCTAAGATCATGTGGTACTTGGATGATCCTGAGCTGACCACCATCA 1864
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 351 LeuValProLeuTyrPheValAlaAlaAlaGluAlaAargLysHisValLysVa 367
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1865 TTGGTCCCCTGACTACTTCGTGGCAGCGGAAGCACGTAAGCACGTCAAGGT 1914
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 367 lValLeuSerGlyGluGlyAlaAspPheGluLeuPheGlyGlyTyrThrIleT 384
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1915 TGTGCTGTCTGGCGAGGGCGCAGATGAGCTGTTCGGTGGATACACCATTT 1964
     |||||||||||||||||||||||||||||||||||||||||||||||||
 384 yrLysGluProLeuSerLeuAlaProPheGluLysLysIleProSerProLeu 400
     |||||||||||||||||||||||||||||||||||||||||||||||||
1965 ACAAAGAGCCGCTATCCGCTTGCTCCATTTGAGAAGATCCCTTCCCCACTA 2014
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 401 ArgLysGlyLeuGlyLysLeuSerLysValLeuProAspGlyMetLysGl 417
     ||||||||||||||||||||||||||||||||||||||||||||||||
2015 CGTAAAGCCTGGGAAAGCTCAGCAAGTTCTGCCACGACGGCATGAAGGG 2064
     |||||||||||||||||||||||||||||||||||||||||||||||
 417 yLysSerLeuLeuGluArgGlySerMetThrMetGluGluArgTyrTyrG 434
     |||||||||||||||||||||||||||||||||||||||||||||||
2065 CAAGTCCCTTCTTGAGCGTGGCTGGCTCCATGACCATGGAAGAGCGCTACTACG 2114
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 434 lyAsnAlaArgSerPheAsnPheGluGlnMetGlnArgValIleProTrp 450
     ||||||||||||||||||||||||||||||||||||||||||||||||
2115 GCAACGAGCTCGCTCCTTCAATTTCGAGCAGATGCAACCGGTTATTCCATGG 2164
     |||||||||||||||||||||||||||||||||||||||||||||||||
 451 AlaLysArgGluTrpAspHisArgGluValThrAlaProIleTyrAlaGl 467
     |||||||||||||||||||||||||||||||||||||||||||||||
2165 GCAAAGCGCGGAATGGGACCACCCGGAAGTCACTGCCACCGATCTACGCACA 2214
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 467 nSerArgAsnPheAspProValAlaArgMetGlnHisLeuAspLeuPheT 484
     ||||||||||||||||||||||||||||||||||||||||||||||||
2215 ATCCCGCAACTTTGATCCAGTAGCCCCGCATGCAACACCTGGATCTGTTCA 2264
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 484 hrTrpMetCysArgGlyAspIleLeuValLysAlaAspLysIleAsnMetAla 500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2265 CCTGGATCGCCGGCGACAATCCTGTCAAGGCTGACAAGATCAACATGGCG 2314
     |||||||||||||||||||||||||||||||||||||||||||||||||
 501 AsnSerLeuGluLeuArgValProPheLeuAspLysLysGluGluValPheLysVa 517
     |||||||||||||||||||||||||||||||||||||||||||||||||||
2315 AACTCCCTTGAGCTGCGAGTTCCATTCTTGGATAAGGAAGTTTTCAAGGT 2364
     |||||||||||||||||||||||||||||||||||||||||||||||||
 517 lAlaGluThrIleProTyrAspLeuLysIleAlaAsnGlyThrThrLysT 534
     |||||||||||||||||||||||||||||||||||||||||||||||||
2365 TGCAGAGACCATTCCTTACGATCTGAAGATTGCCAACGGTACCACCAAGT 2414
     |||||||||||||||||||||||||||||||||||||||||||||||||
 534 yrAlaLeuArgArgAlaLeuGluGluGlnIleLeuValProHisValLeuHis 550
     |||||||||||||||||||||||||||||||||||||||||||||||||
2415 ACGCGCTGCGCAGGGCACTCGAGGAGCAGATTGTTCCGCCTCACGTTTTGCAC 2464
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 551 ArgLysLysLeuGlyPheProValProMetArgHisTrpLeuAlaGlyAs 567
     ||||||||||||||||||||||||||||||||||||||||||||||||
2465 CGCAAGAAGCTGGGCTTCCCTGTTCCCATGCGCCACTGGCTTGCCGGCGA 2514
     |||||||||||||||||||||||||||||||||||||||||||||||||
 567 pGluLeuPheGlyTrpAlaGlnAspThrIleLysGluSerGlyThrGluA 584
     |||||||||||||||||||||||||||||||||||||||||||||||||
2515 TGAGCTGTTCGGTTGGGCGCAGGACACCATTAAGGAATCCGGTACTGAAG 2564
     |||||||||||||||||||||||||||||||||||||||||||||||||
 584 spIlePheAsnLysGlnAlaValLeuAspMetLeuAsnGluHisArgArgAsp 600
     |||||||||||||||||||||||||||||||||||||||||||||||||
```

```
2565 ATATCTTCAACAAGCAGGTGCTGTCTGGATATGCTGAACGAGCACCGCGAT 2614
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 601 GlyValSerAspHisSerArgArgLeuTrpThrValLeuSerPheMetVa 617
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2615 GGCGTGTCAGATCATTCCCGTCGACTGTGACTGTTCTGTCATTTATGGT 2664
     ||||||||||||||||||||||||||||||||||||||||||||||||
 617 lTrpHisGlyValPheValGluAsnArgIleAspProGlnIleGluAspA 634
     ||||||||||||||||||||||||||||||||||||||||||||||||
2665 GTGGCACGGACGCATTTTCTGTGAAAACCGCCATTGATCCACAGATTGAGGACC 2714
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 634 AspA
2715 GCTCCTACCCGGTCGAGCTT 2734
     ||||||||||||||||||||
 634 rgSerTyrProValGluLeu 640
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH67371

seq_documentation_block:
ID   AAH67371 standard; DNA; 1920 BP.
XX
AC   AAH67371;
XX
DT   26-SEP-2001 (first entry)
XX
DE   C glutamicum coding sequence fragment SEQ ID NO: 2406.
XX
KW   Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW   organic acid synthesis; ds.
XX
OS   Corynebacterium glutamicum.
XX
PN   EP1108790-A2.
XX
PD   20-JUN-2001
XX
PF   18-DEC-2000; 2000EP-0127688.
XX
PR   16-DEC-1999; 99JP-0377484.
PR   07-APR-2000; 2000JP-0159162.
PR   03-AUG-2000; 2000JP-0280988.
XX
PA   (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI   Nakagawa S,   Mizoguchi H,   Ando S,   Hayashi M,   Ochiai K,   Yokoi H;
PI   Tateishi N,   Senoh A,   Ikeda M,   Ozaki A;
XX
DR   WPI; 2001-376931/40.
DR   P-PSDB; AAG92152.
XX
PT   Novel polynucleotides derived from Coryneform bacteria, for identifying
PT   mutation point of a gene, measuring expression of a gene, analysing
PT   expression profile or pattern of a gene and identifying homologous gene
PT   -
XX
PS   Claim 8; SEQ ID NO: 2406; 246pp + Sequence Listing; English.
XX
CC   The present invention provides a number of nucleotide and protein
CC   sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC   are useful for identifying the mutation point of a gene derived from a
CC   mutant of coryneform bacterium, measuring expression amount and
CC   analysing the expression profile or expression pattern of a gene derived
CC   from Coryneform bacterium, and identifying a homologue of a gene derived
CC   from coryneform bacterium. Coryneform bacteria are useful for producing
CC   amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC   particularly L-lysine. The present sequence is a nucleic acid described
CC   in the exemplification of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the
CC   European Patent Office.
XX
SQ   Sequence 1920 BP; 422 A; 534 C; 528 G; 436 T; 0 other;

alignment_scores:

```
                 Quality: 3366.00              Length:        640
                 Ratio:      5.268              Gaps:            0
Percent Similarity: 99.844         Percent Identity: 99.688

alignment_block:
US-09-786-474-2  .  x  AAH67371

Align seg 1/1   to: AAH67371   from:  1   to: 1920

   1 MetCysGlyLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPh  17
     |||||||||||||||||||||||||||||||||||||||||||||||
   1 ATGTGCGGCCTTGGCATATTGACCGCAAATGGGAACGCTGAAGCATT  50

  17 eValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA  34
     |||||||||||||||||||||||||||||||||||||||||||||||||
  51 CGTTCCTGCACTCGAGCGGCGGCCTTGCCATGCATGCGCCACCGTGGTCCTG  100

  34 spAspAlaGlyThrTrpHisAspAlaAlaPheGlyPheAsnArg  50
     |||||||||||||||||||||||||||||||||||||||||||||
 101 ACGATGGCGGCACTTGGCATGGACGCCGATGCAGCGTTGGATTCAACCGC  150

  51 LeuSerIleIleAspIleAlaHisSerHisGlnProLeuArgTrpGlyPr  67
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 151 CTCTCCATCATTGATATTGCACACTCCCACCAACCTGCCTTGGGGACC  200

  67 oAlaAspGluProArgTyrAlaMetThrPheAsnGlyGluIleTyrA  84
     ||||||||||||||||||||||||||||||||||||||||||||||||
 201 TGCGGATGAACCCGACCGCTACGCAATGACTTTCAACGGTGAGATCTACA  250

  84 snTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyrPheAsn  100
     ||||||||||||||||||||||||||||||||||| ::::|||||
 251 ACTACGTTGAGCTGCGTGAAAGAGCTCTCGGATTTGGGATATGCCTTTAAT  300

 101 ThrSerGlyAspThrGluHisGlyThrValPheSerSerGluLysGlyl  117
     |||||||||||||||||||||||||||||||||||||||||||||||||
 301 ACTTCTGGCGATGGCGAGCCATGGTCGGTTTCCGTTTCCACCACTGGGGCGA  350

 117 uSerValValGluHisLeuArgGlyMetPheGlyIleAlaIleTrpAspT  134
     |||||||||||||||||||||||||||||||||||||||||||||||||
 351 GTCCGTGGTCGAGCATCTCCGCGGAATGTTCGGCATTGCCATTTGGGATA  400

 134 hrLysGluLysSerLeuPheLeuAlaArgAspGlnIleGluLysLysPro  150
     |||||||||||||||||||||||||||||||||||||||||||||||||
 401 CAAAGGAAAAGTCGCTTTTCCTTGCGCGTGATCAGTTCGGCATTAAGCCA  450

 151 LeuPheTyrAlaThrThrGluHisGlyThrValPheSerSerGluLysLy  167
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 451 CTGTTCTACGCGACCACCGAGCATGGCACCGTCTTCTCCCTCAGAGAAGAA  500

 167 sThrIleLeuGluMetAlaGluHisGlyMetAsnLeuAspLeuGlyLeuL  184
     |||||||||||||||||||||||||||||||||||||||||||||||||
 501 GACCATCTTGGAGATGGCCGAGGAGATGAATCTAGATCTCGGGCCTTGATA  550

 184 ysArgThrIleGluHisTyrValAspLeuGlnThrTyrValProGluProAsp  200
     ||||||||||||||||||||||||||||||||||||||||||||||
 551 AGCGCACCATTGAGCACTACGTGGACTTGCAGTACTGCGCCCAGCCAGAT  600

 201 ThrLeuHisAlaGlnIleSerArgLeuGluSerGlyCysThrAlaThrVa  217
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 601 ACCCTTCACGCGCAGATTTCCCGCTTGGAGTCAGGCTGCACCGCCAACAGT  650

 217 lArgProGlyGlyLysLeuGluGlnLysLysArgTyrPheLysProGlnPheP  234
     |||||||||||||||||||||||||||||||||||||||||||||||||
 651 TCGTCCGGCGGCGGCAAGCTGGAACAGAAGCGTTACTTCAAGCCTCAGTTCC  700

 234 roValGlnLysValValLysGlyValLysGluLysAspLeuPheAspArgIle  250
     ||||||||||||||||||||||||||||||||||||||||||||||
 701 CGGTACAGAGAGTCGTAAAGGGTAAGGAGCAGGACCTCTTCGATCGCATT  750

 251 AlaGlnValLeuGluAspSerSerValGluLysHisMetArgAlaAspValTh  267
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 751 GCCCAGGTGTTGGAGGATAGCTCGGTCGAAAAGCATATGCGTCCCGACGTGAC  800
```

```
 267 rValGlySerPheLeuSerGlyIleGlyIleAspSerThrAlaIleAlaProL  284
     ||||||||||||||||||||||||||||||||||||||||||||||
 801 CCTAGGCTGGTTGCTTTCCGGCGGGGCCAATTGACTCAACCGCAATTGCGGCGC  850

 284 euAlaLysArgHisAsnProAspLeuLeuThrPheThrThrGlyPheGlu  300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 851 TTGCAAAGCGCCACAACCCTGACCTGCTTCACCTTCACCACCGGTTCGAG  900

 301 ArgGluGlyTyrSerGlyGluValAspValAlaGluSerAlaAlaAlaIl  317
     |||||||||||||||||||||||||||||||||||||||||||||||||
 901 CGTGAAGGCTACTCGGAGGTCGATGTGGCTGCGGGAGTCCGCCGCTGCGAT  950

 317 eGlyAlaGluHisIleIleLysIleValSerProGluGluTyrAlaAsnA  334
     |||||||||||||||||||||||||||||||||||||||||||||||||
 951 TGGCGCCTGAGCACATCGTGAAGATTGTCTCGCCTGAAGAATACGCCAACG  1000

 334 laIleProLysIleMetThrTrpTyrLeuAspAspProValAlaAspProSer  350
     |||||||||||||||||||||||||||||||||||||||||||||||
1001 CGATTCCTAAGATCATGTGGTACTTGGATGATCCTGAGCTGACCCCATCA  1050

 351 LeuValProLeuTyrPheValAlaAlaAlaGluAlaArgLysHisValLysVa  367
     |||||||||||||||||||||||||||||||||||||||||||||||||
1051 TTGGTCCCGCTGTACTTCGTGGCAGCCGAAGCACGTAAGCACGTCAAGGT  1100

 367 lValLeuSerGlyGluGlyAlaAspGluLeuPheGlyGlyTyrThrIleT  384
     |||||||||||||||||||||||||||||||||||||||||||||||||
1101 TGTGCTGTCTGGCGAGGGCGCAGATGAGCTGTTCGGTGGATACACCATTT  1150

 384 yrLysGluProLeuSerLeuAlaProPheGluLysIleProSerProLeu  400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1151 ACAAGGAGCCGCTATCGCTTGCTCCATTTGAGAAGATCCCTTCCCCACTA  1200

 401 ArgLysGlyLeuGlyLysLeuSerLysValLeuProAspGlyMetLysGl  417
     |||||||||||||||||||||||||||||||||||||||||||||||||
1201 CGTAAAGGCCTGGGAAAGCTCAGCAAGGTTCTGCCAGACGGCATGAAGGGG  1250

 417 yLysSerLeuLeuGluArgGlySerMetThrMetGluGluArgTyrTyrG  434
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1251 CAAGTCCCTTCTTGAGCGTGGCTCCATGACCATGGAAGAGCGCTACTACG  1300

 434 lyAsnAlaArgSerPheAsnPheGluGlnMetGlnArgValIleProTrp  450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1301 GCAACGCTCGCTCCTTCAATTTCGAGCAGATGCAACGCGTTATTCCATGG  1350

 451 AlaLysArgGluTrpAspHisArgGluValThrAlaProIleTyrAlaGl  467
     |||||||||||||||||||||||||||||||||||||||||||||||||
1351 GCAAAGCGCGAATGGGACCACCGCGAAGTCACTGCGCCGATCTACGCACA  1400

 467 nSerArgAsnPheAsnProValAlaArgMetGlnHisLeuAspLeuPheP  484
     |||||||||||||||||||||||||||||||||||||||||||||||||
1401 GTCCCGCAACTTTGATCCAGTAGCCCGCATGCAACACCTGGATCTGTTCA  1450

 484 hrTrpMetArgGlyAspIleLeuValLysAlaAspLysIleAsnMetAla  500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1451 CCTGGATGCGCGGCGACATCCTGGTCAAGGCTGACAAGATCAACATGGCG  1500

 501 AsnSerLeuGluLeuArgValProPheLeuAspLysGluValPheLysVa  517
     |||||||||||||||||||||||||||||||||||||||||||||||||
1501 AACTCCCTTGAGCTGCGAGTTCCATTCTTGGATAAGGAAGTTTTCAAGGT  1550

 517 lAlaGluThrIleProTyrAspLeuLysIleAlaAsnGlyThrThrLysT  534
     |||||||||||||||||||||||||||||||||||||||||||||||||
1551 TGCAGAGACCATTCCTTACGACCTGAAGATTGCCAACGGTACCACCAAGT  1600

 534 yrAlaLeuArgArgAlaLeuGluGlnIleLeuValProProHisValLeuHis  550
     ||||||||||||||||||||||||||||||||||||||||||||||
1601 ACGGCCTGCGCAGGGCACTCGAGCAGATTGTTCCGCCTCACGTTTTGCAC  1650

 551 ArgLysLysLeuGlyPheGlyTrpProValProMetArgHisTrpLeuAlaGlyAs  567
     |||||||||||||||||||||||||||||||||||||||||||||||||
1651 CGCAAGAAGCTGGGCTTCCCTGTTCCCATGCGCCACTGGCTTGCCGGCGA  1700

 567 pGluLeuPheGlyTrpAlaGlnThrIleLysGluSerGlyThrGluA  584
```

SQ   Sequence 349980 BP; 80724 A; 98367 C; 90490 G; 80399 T; 0 other;

alignment_scores:
             Quality: 3366.00   Length: 640
             Ratio: 5.268   Gaps: 0
Percent Similarity: 99.844   Percent Identity: 99.688

alignment_block:
US-09-786-474-2 x AAH68531

Align seg 1/1   to: AAH68531   from: 1   to: 349980

```
        1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPh 17
          ||||||||||||||||||||||||||||||||||||||||||||||||||
   228516 ATGTGCGGCCTTCTTGGCATATTGACTGCAAATGGGAACGCTGAAGCATT 228565

       17 eValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA 34
          ||||||||||||||||||||||||||||||||||||||||||||||||||
   228566 CGTTCCTGCACTCGAGCGGCCTTGCCATGCACCGCCACCGTGCTCCTG 228615

       34 spAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGlyPheAsnArg 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
   228616 ACGATGCCCGCACTTGGCATGACCCGATGCAGCGTTTCGATTCAACCGC 228665

       51 LeuSerIleIleAspIleAlaHisSerHisGlnProLeuArgTrpGlyPr 67
          ||||||||||||||||||||||||||||||||||||||||||||||||||
   228666 CTCTCCATCATTGATATTGCACACTCCCACCACTCGCGTTGGGGACC 228715

       67 oAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyGlyTyrThrPheAsn 84
          |||||||||||||||||||||||||||||||||||||||::||||||
   228716 TGCGGATGAACCCGACCCGCTACGCAATGACTTTCAACGGTGAGATCTACA 228765

       84 snTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyrThrPheAsn 100
          ||||||||||||||||||||||||||||||||||||||::|||||
   228766 ACTACGTTGAGCTGCGTAAAGAGCTCTCCGATTTGGGATATGCCTTTAAT 228815

      101 ThrSerGlyAspGlyGluProIleValValGlyPheHisHisTrpGlyGl 117
          ||||||||||||||||||||||||||||||||||||||||||||||||||
   228816 ACTTCTGGCGATGGCGAGCCAATTGTTGTCGGTTTCCACCACTGGGGCGA 228865

      117 uSerValValGluHisLeuArgGlyMetPheGlyIleAlaIleTrpAspT 134
          ||||||||||||||||||||||||||||||||||||||||||||||||||
   228866 GTCCGTGGTCGAGCATCTCCGGGATGTTCGGCATTGCCATTTGGGATA 228915

      134 hrLysGluLysSerLeuPheLeuAlaArgAspGlnPheGlyIleLeuLysPro 150
          ||||||||||||||||||||||||||||||||||||||||||||||||||
   228916 CAAAGGAAAAGTCGCTTTTCCTTGCCCGTGATCAGTTCCGCATTAAGCCA 228965

      151 LeuPheTyrAlaThrThrGluHisGlyThrValPheSerSerGluLysLy 167
          ||||||||||||||||||||||||||||||||||||||||||||||||||
   228966 CTGTTCTACGCCACCACCGAGCATGGCACCGTGTTCTCCTCAGAGAAGAA 229015

      167 sThrIleLeuGluMetAlaGluGluMetAsnLeuAspLeuGlyLeuAspL 184
          ||||||||||||||||||||||||||||||||||||||||||||||||||
   229016 GACCATCTTGGAGATGGCCGAGGAGATGAATCTAGATCTGGGCCTTGATA 229065

      184 ysArgThrIleGluHisTyrValAspLeuGlnTyrValProGluProAsp 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||
   229066 AGGCCACCATTGAGCACTACGTGGACCTGCAGTACGTGCCCGAGCCAGAT 229115

      201 ThrLeuHisAlaGlnIleSerArgLeuGluSerGlyCysThrAlaThrVa 217
          ||||||||||||||||||||||||||||||||||||||||||||||||||
   229116 ACCCTTCACGCGCAGATTTCCCGCTTGGAGTCAGGCTGCACCGCAACAGT 229165

      217 lArgProGlyGlyLysLeuGluGlnLysArgTyrPheProGlnPheP 234
          ||||||||||||||||||||||||||||||||||||||||||||||
   229166 TCGTCCGGGCGGCAAGCTGGAACAGAAGCGTTACTTCAAGCCTCAGTTCC 229215

      234 roValGlnLysValValLysGluGlnGlyLysGluAspLeuAspPheAspLeuPheAspAspArgIle 250
          ||||||||||||||||||||||||||||||||||||||||||||||||||
   229216 CAGTACGAAGGTCGTAAAGGGTAAGGACGACCTCTTCGATCGCATT 229265
```

```
     1701 TGAGCTGTGCTGGTGGGCGCCAGGACGACACCATCAAGGAATCCGGTACTGAAG 1750
          ||||||||||||||||||||||||||||||||||||||||||||||||||

      584 spIlePheAsnLysGlnAlaValAlaValLeuAspMetLeuAsnGluHisArgAsp 600
          ||||||||||||||||||||||||||||||||||||||||||||||||
     1751 ATATCTTCAACAAGCAGGCTGTGCTGGATATGCTGAACGAGCACCGCGCAT 1800

      601 GlyValSerAspHisSerArgArgLeuTrpProThrValLeuSerPheMetVa 617
          ||||||||||||||||||||||||||||||||||||||||||||||||
     1801 GGCGTGTCAGATCATTCCCGTCGACTGTGGACTGTTCTGTCATTTATGGT 1850

      617 lTrpHisGlyIlePheValGluAsnArgIleAspProGlnIleGluAspA 634
          ||||||||||||||||||||||||||||||||||||||||||||||||
     1851 GTGGCACGGCATTTTGTGGAAAACCGCATTGATCCACAGATTGAGGACC 1900

      634 rgSerTyrProValGluLeu 640
          |||||||||||||||||
     1901 GCTCCTACCCAGTCGAGCTT 1920
```

seq_name: /SIDS1/gcgdata/geneseqn-embl/NA2001A.DAT:AAH68531

documentation_block:
AAH68531 standard; DNA: 349980 BP.

XX
AC   AAH68531;
XX
DT   26-SEP-2001 (first entry)
XX
DE   C glutamicum coding sequence fragment SEQ ID NO: 7066.
XX
KW   Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW   organic acid synthesis; ds.
XX
OS   Corynebacterium glutamicum.
XX
PN   EP1108790-A2.
XX
PD   20-JUN-2001.
XX
PF   18-DEC-2000;2000EP-0127688.
XX
PR   16-DEC-1999;   99JP-0377484.
PR   07-APR-2000;   2000JP-0159162.
PR   03-AUG-2000;   2000JP-0280988.
XX
PA   (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI   Nakagawa S,   Mizoguchi H,   Ando S,   Hayashi M,   Ochiai K,   Yokoi H;
PI   Tateishi N,   Senoh A,   Ikeda M,   Ozaki A;
XX
PT   WPI; 2001-376931/40.
XX
PT   Novel polynucleotides derived from Coryneform bacteria, for identifying
PT   mutation point of a gene, measuring expression of a gene, analysing
PT   expression profile or pattern of a gene and identifying homologous gene
PT   -
XX
PS   Disclosure; SEQ ID NO: 7066; 246pp + Sequence Listing; English.
XX
CC   The present invention provides a number of nucleotide and protein
CC   sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC   are useful for identifying the mutation point of a gene derived from a
CC   mutant of coryneform bacterium, measuring expression amount and
CC   analysing the expression profile or expression pattern of a gene derived
CC   from Coryneform bacterium, and identifying a homologue of a gene derived
CC   from coryneform bacterium. Coryneform bacteria are useful for producing
CC   amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC   particularly L-lysine. The present sequence is a nucleic acid described
CC   in the exemplification of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the
CC   European Patent Office.
XX

```
251 AlaGlnValLeuGluAspSerValGluLysHisMetArgAlaAspValTh 267
    |||||||||||||||||||||||||||||||||||||||||||||||||
229266 GCCCAGGTGTTGGAGGATAGCCGTGAAAAGCATATGCGTGCCGACGTGAC 229315

267 rValGlySerPheLeuSerGlyLeuValIleAspSerThrAlaIleAlaProL 284
    |||||||||||||||||||||||||||||||||||||||||||||||||||||
229316 CGTAGGCGCTCGTTCCTTTCCGGCGGCATTGACTCAACCGCAATTGCGCCGC 229365

284 euAlaLysArgHisAsnProAspLeuThrPheThrThrGlyPheGlu 300
    ||||||||||||||||||||||||||||||||||||||||||||||||
229366 TTGCAAAGGCGCCACAACCCTGACCTTCACCACCGGTTTCGAG 229415

301 ArgGluGlyTyrSerGluValAlaAspValAlaAlaGluSerAlaAlaAlaIl 317
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
229416 CGTGAAGGCTACTCGGAGGTCGCAGATGGCTGCGGAGTCCGCCGCTGCGAT 229465

317 eGlyAlaGluHisIleValLysIleValSerProGluGluValTyrAlaAsnA 334
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
229466 TGGCGCTGAGCACATCGTGAAGATTGTCTCCCCTGAGGAATACGCCAACG 229515

334 laIleProLysIleMetTrpTyrLeuAspAspProValAlaAspProSer 350
    |||||||||||||||||||||||||||||||||||||||||||||||||
229516 CGATTCCTAAGATCATGTGGTACTTGGATGATCCTGTAGCTGACCCATCA 229565

351 LeuValProLeuTyrPheValAlaAlaAlaArgLysHisValLysVa 367
    ||||||||||||||||||||||||||||||||||||||||||||||||
229566 TTGGTCCCGCTGTACTTCGTGGCAGCGGCCGCTAAGCACGTCAAGGT 229615

367 lValLeuSerGlyLeuGlyAlaAspGluPheGlyGlyTyrThrIleT 384
    |||||||||||||||||||||||||||||||||||||||||||||||
229616 TGTGCTGTCTGGCGAGGGCGCAGATGAGCTTGGTGGATACACCATTT 229665

384 yrLysGluProLeuSerLeuAlaProPheGluLysIleProSerProLeu 400
    |||||||||||||||||||||||||||||||||||||||||||||||||
229666 ACAAGGAGCCCCTATCGCTTGCTCCATTTGAGAAGATCCCTTCCCCACTA 229715

401 ArgLysGlyLeuGlyLysLysSerLysValLeuProLysGlyMetLysGl 417
    ||||||||||||||||||||||||||||||||||||||||||||||||||
229716 CGTAAAGGGCTGGGAAAGCTCAGCAAGGTTCTGCCAAAGGGCATGAAGGG 229765

417 yLysSerLeuLeuGluArgGlySerMetThrMetGluGluArgTyrTyrG 434
    |||||||||||||||||||||||||||||||||||||||||||||||||
229766 CAAGTCCCTTCTTGAGCGTGGCTCCATGACCATGGAAGAGCGCTACTACG 229815

434 lyAsnAlaArgSerPheAsnProGluAsnPheGluGlnMetGlnArgValIleProTrp 450
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
229816 GCAACGCTCGCTCCTTCAATTTCGAGCAGATTCGACAGCCGGTTATTCCATGG 229865

451 AlaLysArgGluTrpAspHisArgGluValThrAlaProIleTyrArgAlaGl 467
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
229866 GCAAAGCGCGAATGGGACCACCGCGAAGTCACTGCCCCGATCTACCGCACA 229915

467 nSerArgAsnPheAspProValAlaArgMetGlnHisLeuAspLeuPheT 484
    |||||||||||||||||||||||||||||||||||||||||||||||||
229916 GTCCCCAACTTTGATCCGGTAGCCCGCATGCAACACCTGGATCTGTTCA 229965

484 hrTrpMetArgGlyAspIleLeuValLysAlaAspLysIleAsnMetAla 500
    ||||||||||||||||||||||||||||||||||||||||||||||||||
229966 CCTGGATGCGCGGCGACATCCTGGTCAAGGCTGACAAGATCAACATGGCG 230015

500 AsnSerLeuGluLeuArgValProPheLeuAspProPheLysGlyValPheLysVa 517
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
230016 AACTCCCTTGAGCTGCGAGTTCCATTCTTGGATCCATTCAAGGGTGTCAAGGT 230065

517 lAlaGluThrIlePheTyrAspTyrArgLeuLysIleAlaAsnGlyThrThrLysS 534
    |||||||||||||||||||||||||||||||||||||||||||||||||||||
230066 TGCAGAGACCATTCCTTACGACCTGCGAGATTGCCAACGGTACCACCAAGT 230115

534 yrArgAlaLeuArgAlaLeuGluGlnIleLeuValProHisValLeuHis 550
    ||||||||||||||||||||||||||||||||||||||||||||||||||
230116 ACGCGCTGCCGCAGGGCACTCGAGCACTGTTCCCGCCTCACGTTTTGCAC 230165

551 ArgLysLysLeuGlyPheValProMetArgHisTrpLeuAlaGlyAs 567
    ||||||||||||||||||||||||||||||||||||||||||||||||
```
```
    |||||||||||||||||||||||||||||||||||||||||||||||
230166 CGCAAGAAGCTGGGCTTCCCTGTTCCCATGCGCCACTGGCTTGCCGGCGA 230215

567 pGluLeuPheGlyTrpProAlaGlnAspThrIleLysSerGlyThrGluA 584
    |||||||||||||||||||||||||||||||||||||||||||||||||
230216 TGAGCTGTTCGGTTGGCCGGCGCAGGACACCATCAAGGAATCCGGTACTGAAG 230265

584 spIlePheAsnLysGlnAlaValAlaValLeuAspMetLeuAsnGluHisArgAsp 600
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
230266 ATATCTTCAACAAGCAGGCTGTGTCTGTGATATGCTGGATAGCTGAACGAGCACCGCGAT 230315

601 GlyValSerAspHisSerArgArgLeuTrpThrValLeuSerPheMetVa 617
    |||||||||||||||||||||||||||||||||||||||||||||||||
230316 GGCGTGTCAGATCATTCCCGTGGACTGTGGACTGTTCTGTCATTTATGGT 230365

617 lTrpHisGlyIlePheGluGluAsnArgIleAspProGlnIleGluAspA 634
    |||||||||||||||||||||||||||||||||||||||||||||||||
230366 GTGGCACGGACCATTTTGTGGAAAAACCGCATTGATCCACAGATTGAGGACC 230415

634 rgSerTyrProValGluLeu 640
    ||||||||||||||||||||
230416 GCTCCTACCCAGTCGAGCTT 230435
```

seq_name: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:AAF71807

Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:109.

```
seq_documentation_block:
ID    AAF71807 standard; DNA; 1962 BP.
XX
AC    AAF71807;
XX
DT    30-APR-2001  (first entry)
XX
DE    Corynebacterium glutamicum MP protein.
XX
KW    Corynebacterium glutamicum; metabolic pathway protein; MP protein;
KW    fine chemical production; microorganism; organic acid; nucleoside;
KW    nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
KW    lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
KW    carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.
XX
OS    Corynebacterium glutamicum.
XX
PN    WO200100843-A2.
XX
PD    04-JAN-2001.
XX
PF    23-JUN-2000   2000WO-IB00923.
XX
PR    25-JUN-1999;  99US-0141031.
PR    01-JUL-1999;  99DE-1030476.
PR    02-JUL-1999;  99US-0142101.
PR    08-JUL-1999;  99DE-1031415.
PR    08-JUL-1999;  99DE-1031418.
PR    08-JUL-1999;  99DE-1031419.
PR    08-JUL-1999;  99DE-1031420.
PR    08-JUL-1999;  99DE-1031424.
PR    08-JUL-1999;  99DE-1031428.
PR    08-JUL-1999;  99DE-1031434.
PR    08-JUL-1999;  99DE-1031435.
PR    08-JUL-1999;  99DE-1031443.
PR    08-JUL-1999;  99DE-1031453.
PR    08-JUL-1999;  99DE-1031457.
PR    08-JUL-1999;  99DE-1031465.
PR    08-JUL-1999;  99DE-1031478.
PR    08-JUL-1999;  99DE-1031510.
PR    08-JUL-1999;  99DE-1031541.
PR    08-JUL-1999;  99DE-1031573.
PR    08-JUL-1999;  99DE-1031592.
PR    08-JUL-1999;  99DE-1031632.
PR    08-JUL-1999;  99DE-1031634.
PR    08-JUL-1999;  99DE-1031636.
PR    09-JUL-1999;  99DE-1032125.
PR    09-JUL-1999;  99DE-1032126.
```

PR   09-JUL-1999;   99DE-1032130.
PR   09-JUL-1999;   99DE-1032186.
PR   09-JUL-1999;   99DE-1032206.
PR   09-JUL-1999;   99DE-1032227.
PR   09-JUL-1999;   99DE-1032228.
PR   09-JUL-1999;   99DE-1032229.
PR   09-JUL-1999;   99DE-1032230.
PR   14-JUL-1999;   99DE-1032922.
PR   14-JUL-1999;   99DE-1032926.
PR   14-JUL-1999;   99DE-1032928.
PR   14-JUL-1999;   99DE-1033004.
PR   14-JUL-1999;   99DE-1033005.
PR   14-JUL-1999;   99DE-1033006.
PR   12-AUG-1999;   99US-0148613.
PR   27-AUG-1999;   99DE-1040764.
PR   27-AUG-1999;   99DE-1040765.
PR   27-AUG-1999;   99DE-1040766.
PR   27-AUG-1999;   99DE-1040832.
PR   31-AUG-1999;   99DE-1041378.
PR   31-AUG-1999;   99DE-1041379.
PR   31-AUG-1999;   99DE-1041380.
PR   31-AUG-1999;   99DE-1041394.
PR   31-AUG-1999;   99DE-1041396.
PR   03-SEP-1999;   99DE-1042076.
PR   03-SEP-1999;   99DE-1042077.
PR   03-SEP-1999;   99DE-1042079.
PR   03-SEP-1999;   99DE-1042086.
PR   03-SEP-1999;   99DE-1042087.
PR   03-SEP-1999;   99DE-1042088.
PR   03-SEP-1999;   99DE-1042095.
PR   03-SEP-1999;   99DE-1042124.
PR   03-SEP-1999;   99DE-1042129.
PR   09-MAR-2000;   2000US-0187970.
XX
PA   (BADI) BASF AG.
XX
PI   Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
DR   WPI; 2001-137957/14.
DR   P-PSDB; AAB79688.
XX
PT   Nucleic acids from Corynebacterium glutamicum encoding metabolic
PT   pathway proteins, useful for producing fine chemicals in
PT   microorganisms, including organic acids, nonproteinogenic amino acids,
PT   and purine and pyrimidine bases -
XX
PS   Claim 3; Page 324-327; 1737pp; English.
XX
CC   AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
CC   pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
CC   MP nucleic acids are useful for the production of fine chemicals
CC   in microorganisms, including organic acids, nonproteinogenic amino
CC   acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
CC   saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
CC   compounds, vitamins, cofactors, polyketides and enzymes.
XX
SQ   Sequence 1962 BP; 435 A; 544 C; 534 G; 449 T; 0 other;

alignment_scores:
      Quality: 3360.00   Length: 640
        Ratio: 5.266   Gaps: 0
Percent Similarity: 99.688   Percent Identity: 99.531

alignment_block:
US-09-786-474-2 x AAF71807   ..

Align seg 1/1 to: AAF71807 from: 1 to: 1962

```
   1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPh  17
     |||||||||||||||||||||||||||||||||||||||||||||||||
  20 ATGTGCGGCCTTCTTGGCATATTGACTGCAAATGGGAACGCTGAAGCATT  69
```

```
  17 eValProAlaAlaGluArgAlaAlaLeuProCysMetArgHisArgGlyProA  34
     |||||||||||||||||||||||||||||||||||||||||||||||||||||
  70 CGTTCCTGACTCGACGCGGGCCTTGCCATGCATGCGACCACCGTGGTCCTG  119
```

```
  34 spAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGlyPheAsnArg  50
     |||||||||||||||||||||||||||||||||||||||||||||||||
 120 ACGATGCCGGCACTTGGCATGACGCCGATTGCAGCGTTTGGATTCAACCGC  169
```

```
  51 LeuSerIleIleAspIleAlaHisSerHisGlnProLeuArgTrpGlyPr  67
     ||||||||||||||||||||||||||||||||||||||||||||||||
 170 CTCTCCATCATTGATATTGCACACTCCCACCAACCACTGCGTTGGGGACC  219
```

```
  67 oAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyGluIleTyrA  84
     |||||||||||||||||||||||||||||||||||||||||||||||||
 220 TGCCGGATGAACCCGACCCGCTACGCCAATGACTTTCAACGGTGAGATCTACA  269
```

```
  84 snTyrValGluLeuArgLysGluLysGlyTyrThrPheAsn  100
     |||||||||||||||||||||||||||||||||||:::||||
 270 ACTACGTTGAGCTGCGTAAAGAGCTCCGATTTGGGATATGCCTTTAAT  319
```

```
 101 ThrSerGlyAspGlyGluProIleValValGlyPheHisHisTrpGlyGl  117
     |||||||||||||||||||||||||||||||||||||||||||||||||
 320 ACTTCTGGCAGTGGCGAGCCAATTGTTGTCGGTTTCCACCACTGGGGCGA  369
```

```
 117 uSerValValGluHisLeuArgGlyMetPheGlyIleAlaIleTrpAspT  134
     |||||||||||||||||||||||||||||||||||||||||||||||||
 370 GTCCGTGGTCGAGCATCTCCGGCAATGTTCGGCATTGCCATTTGGGATA  419
```

```
 134 hrLysGluLysSerLeuPheLeuAlaArgAspGlnPheGlyIleLysPro  150
     |||||||||||||||||||||||||||||||||||||||||||||||||
 420 CAAAGGAAAAGTCGCTTTTCCTTGCGCGTGATCAGTTCGGCATTAAGCCA  469
```

```
 151 LeuPheTyrAlaThrArgGluHisGlyThrValPheSerGluGluLysLy  167
     |||||||||||||||||||||||||||||||||||||||||||||||||
 470 CTGTTCTACGCAACCACCGAGCATGGCACCGTGTTCTCCTCAGAGAAGAA  519
```

```
 167 sThrIleLeuGluMetAlaGluGluMetAsnLeuAspLeuGlyLeuAspL  184
     |||||||||||||||||||||||||  |||||||||||||||||||||
 520 GACCATCTTGGAGATGGCCGAGGAGATGAATCTAGATCTGGGCCTGATA  569
```

```
 184 ysArgThrIleGluHisTyrValAspLeuGlnThrValValProGluProAsp  200
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 570 AGGCGGCACCATTGAGCACTACGTGGACTTGCAGACGTGCCCGAGCCAGAT  619
```

```
 201 ThrLeuHisAlaGlnIleIleSerArgLeuGluSerGlyCysThrAlaThrVa  217
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 620 ACCCTTCACGCGCAGATTTCCCGCTTGGAGTCAGGCTGCACCGCAACAGT  669
```

```
 217 lArgProGlyGlyLeuGluGlnLnLysArgTyrPheLysProGlnPheP  234
     ||||||||||||||||||||||||||||||||||||||||||||||||
 670 TCGTCCGGGGCGGCAAGCTGGAACAGAAGCCTTACTTCAAGCCTCAGTTCC  719
```

```
 234 roValGlnLysValValGlyGlyLysGluGlnAspLeuPheAspArgIle  250
     ||||||||||||||||||||||||||||||||||||||||||||||||
 720 CAGTACAGAAGGTCGTAAAGGGTAAGGAGCAGGACCTCTTCGATCGCATT  769
```

```
 251 AlaGlnAsnValLeuGluAspSerValGluLysHisMetArgAlaAspValTh  267
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 770 GCCCAGGTGTTGGAGGATAGCCTCGAAAAGCATATGCGTGCCGACGTGAC  819
```

```
 267 rValGlySerPheLeuSerGlyIleGlyIleAspSerThrAlaIleAlaProL  284
     |||||||||||||||    |||||||||||||||||||||||||||||||
 820 CGTAGGCTCGTTCTTTGCGGGCATTGACTCAACGCAATTGCGGGCGCGC  869
```

```
 284 euAlaLysArgHisAsnProAspLeuThrPheThrThrGlyPheGluGlu  300
     |||||||||||||||||||||||||||||||||||||||||||||||||
 870 TTGCAAAGGCGCCACAACCCTGACCTTCACCTTCACCACCGGTTTCGAG  919
```

```
 301 ArgGluGlyTyrSerGluValAspValAspValAlaAlaLeuSerAlaAlaAlaI  317
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 920 CGTGAAGGCTACTCGGAGGTCGATGTGGCTGCGGAGTCGCCGCTGCGAT  969
```

```
 317 eGlyAlaGluHisIleValLeuValSerProGluGluTyrAlaAsnA  334
     ||||||||||||||||||||||||||||||||||||||||||||||||
```

```
        |||||||||||||||||||||||||||||||||||||||||||||||||
   970  TGGCCTGAGCACATCGTGAAGATTGTCTCCCCTGAGGAATACGCCAACG  1019
   334  laIleProLysIleMetTrpTyrLeuAspAspProValAlaAspProSer  350
  1020  CGATTCCTAAGATCATGTGGTACTTGGATGATCCTGACCCCATCA  1069

        |||||||||||||||||||||||||||||||||||||||||||||||||
   351  LeuValProLeuTyrPheValAlaAlaAlaGluAlaArgLysHisValLysVa  367
  1070  TTGGTCCCGCTGTACTTCGTGGCAGCGGAAGCACGTAAGCACGTCAAGGT  1119

   367  lValLeuSerGlyGlyAspGluLeuPheGlyGlyTyrThrIleT  384
  1120  TGTGCTGTCTGGCGAGGGCGACGATGAGCTGTTCGGTGGATACACCATTT  1169

   384  yrLysGluProLeuSerLeuAlaProGluLysIleProSerProLeu  400
  1170  ACAAGGAGCCCCTATCGCTTGCTCCTGAGAAGATCCCCTCCCCACTA  1219

   401  ArgLysGlyLeuGlyLysLeuSerLysValLeuProAspLysMetLysGl  417
  1220  CGTAAGGGCTGGGAAAGCTCAGCAAGGTTCTGCCAGACGGCATGAAGGG  1269

   417  yLysSerLeuLeuGluGlySerMetThrMetGluGluLysArgTyrTyrG  434
  1270  CAAGTCCCTCCTTCTTGAGCGGTGCTCCATGACCATGGAAGAGCGCTACTACG  1319

   434  lyAsnAlaArgSerPheAsnPheGluGlnMetGlnArgValIleProTrp  450
  1320  GCAACGCTCGCTCCTTCAATTCGAGCAGATGCAACGCGTTATTCCATGG  1369

   451  AlaLysArgGluTrpAspHisArgAspHisArgGluValThrAlaProIleTyrAlaGl  467
  1370  GCAAAGCGCGAATGGGACCACCGGGAAGTCACTGCGCCGATCTACGCACA  1419

   467  nSerArgAsnPheAspProValAlaArgMetGlnHisLeuAspLeuPheT  484
  1420  GTCCCGCAACTTTGATCCAGTAGCCCGGATGCAACACCTGGATCTGTTCA  1469

   484  hrTrpMetArgGlyAspIleLysAspLysAlaAspLysIleAsnMetAla  500
  1470  CCTGGATGCGCGGCGACATCGTGGTCAAGGCTGACAGATCAACATGGCG  1519

   501  AsnSerLeuGluLeuArgValProPheLeuAspLysValPheLysVa  517
  1520  AACTCCCTTGAGCTGCGAGTTCCATTCTTGGATAAGGAAGTTTTCAAGGT  1569

   517  lAlaGluThrIleProTyrAspLeuLysLysIleAlaAlaAsnGlyThrThrLysT  534
  1570  TGCAGAGACCATTCCTTACGACCTGAAGAAGATTGCCAACGGTACCACCAAGT  1619

   534  yrAlaLeuArgArgAlaLeuGluGluValIleProHisValLeuHis  550
  1620  ACGCGCTGCGCAGGGCACTCGAGGAGGTTGTTCCGCCTCACGTTTGCAC  1669

   551  ArgLysLysGlyLeuGlyPheProValProMetArgHisTrpLeuAlaGlyAs  567
  1670  CGCAAGAAGGGCTGGGGCTTCCCTGTTCCCATGCGCCACTGGCTTGCCGGCGA  1719

   567  pGluLeuPheGlyTrpTrpAspGlnAspThrIleLysSerGlyThrGluA  584
  1720  TGAGCTGTTCGGTTGGTGGGACCAGGACACCATCAAGGAATCCGGTACTGAAG  1769

   584  spIleAsnAlaValLeuAspMetLeuAsnGluHisArgAspAsp  600
  1770  ATATCTTCAACAAGCAGGCTGTCCTGGATATGCTGAACGAGCACCGCGAT  1819

   601  GlyValSerArgAspHisSerArgArgLeuTrpThrValLeuSerPheMetVa  617
  1820  GGCGTGTCAGATCATTCCCGTCGACTGTTCTGTTCATTTATGGT  1869

   617  lTrpHisGlyIleGluPheValGluAsnArgIleAsnAspProGlnIleGluAspAspA  634
```

```
alignment_scores:
        Quality:    633.00           Length:       622
        Ratio:      1.829            Gaps:          20
Percent Similarity: 55.627     Percent Identity: 29.582

alignment_block:
US-09-786-474-2 x AAT61016/rev

Align seg 1/1    to reverse of: AAT61016    from: 1   to: 30001

     1  MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPh   17
 22914  ATGTGCGGAATCGTGGGGGTGGTCGACTACGACGCCCGGCCGAACACCG  22865

    17  eValProAlaLeuArgAlaLeuProCysMetArgHisArgGlyProA   34
```

```
307   alAspValAlaAlaGluSerAlaAlaAlaAlaIleGlyAlaGluHisIle...  322
      ::  .:: |||||||||||:::|||||
21998 CGCCTACGTGCCGAGCTGGCGGGGCACGTGGCCGAGCGGACCACCTGGAG      21949

323   ValLysIleValSerProGluGluTyrAlaAsnAlaIleProLysIleMe    339
      |||  .::: |||||
21948 GTGCTGCTGGACAGCCCCGGC..................                 21928

339   tTrpTyrLeuAspAspProAlaAlaAspProSerLeuVal               352
      |||:.||||||:: :::::::
21927 ........CTGTCGGACCCGGCGGTGCGTTCCGGCGCTGCGCCACCG         21885

353   ........                         ProLeuTyrPheVal        357
                                       |||||||
21884 ACCTGCCGCCCCGGCTACTGGGGCGACATGGCCCTCGCTCTACCTGTTC      21835

358   AlaAlaGluAlaArgLysHisValLysValLeuSerGlyValGluGlyAl    374
      :::|||||||||| |||:||||||||||||||:::|||
21834 TTCCGGCGGTGCGGGAGAACGTCACGGTGGCCGTGTCCGGCGAGGCCGC      21785

374   aAspGluLeuPheGlyGlyTyrThrIleTyrLysGluProLeuSerLeu.    390
      |||||||||||||| :::::::::|||:|||
21784 GGACGAACTTCGGCGGCTACCGTGGTTCCGCAACCCGTTGGCCCTGC        21735

391   ....AlaProPheGluLysProSerProLeuArgLysGlyLeuGly        405
                        |||||||||||
21734 GGGCGGCCACCTTC.                                          21721

406   LysLeuSerLysValLeuProAspGlyMetLysGlyLysLysSerLeuGl    422
                                               |||
21720 ..................CCGTGGCTCAC                            21710

422   uArgGlySerMetThrMetGluGluArgTyrTyrGlyAsnAlaArgSerP    439
      ||||||  ::::::::||||
21709 CCCGGGTTCG................GCCGCTACTTCGGCGGCAGCTCGCTGT  21672

439   heAsnPheGluGlnMetGlnAsnArgValIleProTrpAlaLysArgGluTrp 455
      |||||  ::::||||||||||||::::|||
21671 TCCAGGCGGGTGCTGGACAAGTCGACCTGGACGGCTACCGGCGGCC        21622

456   AspHisArgGluValThrAlaProIleTyrAlaGlnSerArgArgAsnPheAs 472
      ||| :::  :::
21621 CGGTACGCCGAGGCCGTCGCCGAAGTGCCCGGTCTGCCGGGTGAGTCGGC    21572

472   pProValAlaArgMetGln........        HisLeuAspLeuPheThrTrpM 486
      |||||||:::|||
21571 GCCCGAGACGGCGATGCCCGCGAGATCGGTCACCTGAACCTGACCCGCTTCG  21522

486   etArgGlyAspIleLeuValLysAlaAspLysIleAsnMetAlaAsnSer   502
      ::::: .::
21521 CAACGTGCCCTGGGACGATGAAGAGTTCGACGGCCGGAGAGAGCCTGC      21475

503   LeuGluLeuArgValProPheLeuLysProLeuLysLysValAlaGl     519
      ||||||||:::|||||| ||| :::
21474 CTCGAGGTGCGGGTGCCGTTCCTGCCCACCACCGGCTCGGAGTACGTCTT    21425

519   uThrIleProTyrAspLeuLysIleAlaAsnGlyThrThrLysTyrAlaL   536
      |||   :: :::
21424 CAACGTGCCCTGCCCGGGAGATGAAGAGTTCGACGGCCGGAGAGAGCCTGC  21375

536   euArgArgAlaLeuGluGlnIleValProProHisValLeuHisArgLys   552
      |||||  ||| ||||||||||||::::|||||
21374 TGCGGGGCGCGGTGGCCGAACCTGGTCGGTCGACGCGGGTG            21325

553   LysLeuGlyPheProVal                                     561
      |||  ::::|||
21324 AAGACCCCGTACCCGGCACGACCCCAGGATCCCGGGCTACGAGGGGCCCCCCGCCG 21275

561   gHisTrpLeuAlaGly  566
      |||  ::::|||
21274 GGCCGAACTGGCCCGGG  21259
```

```
22864 CCGGGACCTGCTCCAGGCGGATGACCACCATGGCCTGCCGGGGGCCGG   22815

34    spAspAlaGlyThrTrpHisAspAlaAspAlaPheGlyPheAsnArg    50
      |||  ::||||||| :::|||
22814 ACGCCGAGGGCCTCGGTTCGCGCCGGCTCGGTCACCGCGC            22765

51    LeuSerIleIleAspIleAlaHisSerHisGlnProLeuArgTrpGlyPr  67
      |||:||||| ||| |||
22764 CTGTCGGTGATCGACCCCGAGCACGGCCAGCCGATG.......AC      22724

67    oAlaAspGluProAspArg...TyrAlaMetThrAspAsnGlyIleThr  83
      |||:.||||||: :::::
22723 CGCGGAACACGAGGGCCCGCGGCCGGCGATCACCTTCAGCGGCGAGATCT  22674

83    yrAsnTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyrThrPhe  99
      |||||||
22673 ACAACTTCCGTGAGCTGCGCAAGGAACTCACCTCGCACGGCCACCGGTTC  22624

100   AsnThrSerGlyAspGlyGluProIleValGlyPheHisHisTrpGl    116
      :::||||| ||| :::|||||
22623 AGGACGTCTCGGATACCGAGGTCGTCCTGCGGCCGCTACCTCCAGTGGGG  22574

116   yGluSerValValGluHisLeuHisLeuArgGlyMetPheGlyIleAlaIleTrpA 133
      | |:.||||||||||||::|||:::|||
22573 CGCCCGGCTGGTCGAGAGGCTCAACGGAGTGTTCGGCTTCGCCCTCTGGG  22524

133   spThrLysLysSerLeuPheLeuAlaArgAspGlnPheGlyLysIleLys  149
      :::||||| ::|||||||||||:::|||
22523 ACGAGCACCAGCGGAGAACTCCGTGGTCCGGACCCGATGGGCGTCAAG    22474

150   ProLeuPheTyrAlaThrThrGlyHisGlyThrValPheSerSerGluLu  166
      ||||||:::||| ||| :::|||||
22473 CCGCTCTACTACTGGCCGACCGCCCGCCGGGTGCGCCTTCGGCTCGGAGCC 22424

166   sLysThrIleLeuGluMetAlaGluGluMetAsnLeuAspAsnGlyLeuSerGlyLeuA 183
      | :::||| |||
22423 CAAGGGCCGTCCTG.................GCGGACCGGACCCTG       22395

183   spLysArgThr........        IleGluHisTyrValAspLeuGlnTyr  195
      ||| :::
22394 CCCGGGGTCGGTGTGGAGTGCGTGGACATGGTGAAG                22345

196   ValProGluProAspThrLeuHisAlaGlnIleSerArgLeuGluSerGl  212
      |::.|||||
22344 ACCCCGGAGGCCGGCCGTCTTCTCCGGCGCGTCGAG                22309

212   yCysThrAlaThrValArgProGly..........        GlyL     222
      |||||
22308 .........GTCCGTCCGGGTCAGCTCGTCCGGGGGGCCGGCG         22272

222   ysLeuGluGlnIleLysTyrPheLysProGlnPheProValGlnLysVal  238
      :::::|||| :::
22271 GGGTGCGGCCGGCCAGACTGGCGGGCTGGAG                     22240

239   ValLysGlyLysGluGlnMetAspLeu........        PheAspArgIleAlaGl 252
      :::||| ||
22239 GCCCGAGGCGGAGCGACCTGCCCGACGACCATCCGCACCTCCGTGA      22190

252   nValLeuGluAspSerValGluLysHisMetArgValAlaAspValThrValG 269
      :::|| :::|||
22189 CCTGCTCGCCGACACCGTCACCCGTGGTCTCGGACGTCCCCGCTGG      22140

269   lySerPheLeuSerGlyLyIleLeuAspSerThrAlaIleAlaProLeuAla 285
      ||||||
22139 GCACCGTCTCCTCCGGGACTTCGTCACGGCGGCTGCCC              22090

286   LysArgHisAsnProAspLeuLeuThrPheThrThrGlyGlyPhe........  299
      |||  :::
22089 GCCCGCAGCGACGGGCCG.....GTGCGTCCCTTCTCGACTTCGGGGGC   22043

300   ........                         GluArgGluGlyTyrSerGluV  307
                                       |||||
22042 CGCCGACGCGCCTTCGCCGCACGCCGTTCGCGAGGCCGAG.....GCGACG 21999
```

us-09-786-474-2.p2n.rng

seq_name: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1999.DAT:AAX05110

```
seq_documentation_block:
ID   AAX05110 standard; DNA; 30001 BP.
XX
AC   AAX05110;
XX
DT   07-APR-1999 (first entry)
XX
DE   S. aureofaciens DNA in cosmid clones LP2-127 and LP2-128.
XX
KW   Chlortetracycline; tetracycline; biosynthetic; actinomycete;
KW   S. ariseofuscus; S. ambofaciens; antibiotic resistance gene;
KW   S. lividans; S. aureofaciens; cosmid clone; antibiotic; ss.
XX
OS   Streptomyces aureofaciens.
XX
RN   US5866410-A.
RD   02-FEB-1999.
XX
PF   07-JUN-1995;   95US-0474933.
XX
PR   22-SEP-1993;   93US-0125468.
PR   26-JUL-1990;   90US-0558039.
PR   26-JUL-1990;   90US-0558040.
PR   26-JUL-1990;   90US-0821109.
PR   15-JAN-1992;   92US-0821419.
PR   07-JUN-1995;   95US-0474933.
XX
PA   (AMCY ) AMERICAN CYANAMID CO.
XX
PI   Fantini SE,   Lotvin JA,   Ryan MJ,   Strathy N;
XX
DR   WPI; 1999-141936/12.
XX
PT   Two plasmids for cloning the biosynthetic pathways of tetracycline,
PT   chlortetracycline, and their derivatives - comprise an origin of
PT   replication, cohesive end sites and optionally, an antibiotic
PT   resistance gene
XX
PS   Disclosure; Fig 4A-L; 39pp; English.
XX
CC   The invention relates to combination of two plasmids for cloning the
CC   genes encoding the biosynthetic production pathways of chlortetracycline,
CC   tetracycline, or their derivatives. One plasmid contains an origin of
CC   replication, an actinomycete (such as Streptomyces lividans,
CC   S. ariseofuscus, or S. ambofaciens) active antibiotic resistance gene and
CC   three or more tandem cohesive end sites. The second plasmid contains an
CC   actinomycete active origin of replication and three or more tandem
CC   cohesive end sites. The invention can be used to clone a large amount of
CC   genetic material for the heterologous production of the antibiotics,
CC   chlortetracycline, tetracycline and their analogues. A cluster of genes
CC   can be inserted into the plasmid allowing a biosynthetic pathway to be
CC   transferred in its entirety to a heterologous host. The invention
CC   specifically relates to cloning of the entire tetracycline and
CC   chlortetracycline from Streptomyces aureofaciens and its expression in a
CC   heterologous host such as S. lividans. The present sequence represents a
CC   S. aureofaciens DNA sequence from the cosmid clones designated LP2-127
CC   and LP2-128.
XX
SQ   Sequence 30001 BP; 4055 A; 10939 C; 11019 G; 3988 T; 0 other;
```

```
alignment_scores:
     Quality:      633.00        Length:          622
     Ratio:         1.829        Gaps:             20
Percent Similarity:  55.627   Percent Identity: 29.582
```

```
alignment_block:
US-09-786-474-2 x AAX05110/rev     ..

Align seg 1/1 to reverse of: AAX05110   from: 1  to: 30001
```

```
    1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPh   17
      ||||||||||||:::::|||      ::::|        |||||||
22914 ATGTGCGGAATCGTGGGGGTGGTTCGACTACGAGCCCCGGCAACACCG   22865

   17 eValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA   34
      ||||||  |||     :::|    |||||||:|||:|||||||||||
22864 CCGGGACCTGCTCCAGGCGATGACCACCACCATGGCCTGCCCGGGGGCCGG   22815

   34 sPAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGlyPheAsnArg   50
      |||   |||  |||::|:|||  :::|||      ||||||
22814 ACGCCGAGGGCCTCTGTTCGCCCGCCGGGCTCGGTCACCGGCGC   22765

   51 LeuSerIleIleAspIleAlaHisSerHisGlnProLeuArgTrpGlyPr   67
      ||||||:|| ||||::::||:||||      ::|
22764 CTGTCGGTGATCGACCCCGAGCACGGCCGCAGCCGATG........AC   22724

   67 oAlaAspGluProAspArg...TyrAlaMetThrPheAsnGlyGluIleT   83
      ||||||:| ||||::::|||||:||::||||||||||
22723 CGCGGAACACGAGGGCCGCGGCCGATCACCTTCAGCGGCGAGATCT   22674

   83 yrAsnTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyrThrPhe   99
      |||||:| |||||||||||  :::::|| |||    :||   |||
22673 ACAACTTCCGTGAGCTGCCCGAACTCACCTCGCACGGCCACCGGTTC   22624

  100 AsnThrSerGlyAspGlyGluProIleValGlyPheHisHisTrpGl   116
      ||||||:|:|||  |||  |||::|||||||||:::||:||||||:|||
22623 AGGACGTCCTGCGATACCGAGGTCGTCCTGCGGGCTACCTCCAGTGGGG   22574

  116 yGluSerValValGluHisLeuArgGlyMetPheGlyIleAlaIleTrpA   133
      |   ||||||:|||:::||||:||||||||||||:::|||||||||||||
22573 CGCCGGCTGGTCGAGAGGCTCAACGGCGATGTTCGCCTTCGCCGTCTGGG   22524

  133 spThrLysGluLysSerLeuPheLeuAlaLeuAlaArgAspGlnPheGlyIleLeuLys   149
      ||  :::::|||||||:|||::|||||||||||||||||:::||||||
22523 ACGAGCACAGCGAGGAACTCCGTCGTTCCGCGACCGGATGGGCGTCAAG   22474

  150 ProLeuPheTyrAlaThrThrGluHisGlyThrValPheSerGluLy   166
      |||||||||||::|||   |||     :::|||||:|||||:|||||||
22473 CCGCTCTACTACTGGCCCGACCGCGCGTCCCGGCGTCGCGTTCGGCTCGGAGCC   22424

  166 sLysThrIleLeuGluMetAlaGluGluMetAsnLeuAspLeuGlyLeuA   183
      |||   ||||          ||    |||:|:|||
22423 CAAGGCCCGTCCTG................GCCGACCGGACCCTGT   22395

  183 spLysArgThr.........IleGluHisTyrValAspLeuGlnTyr   195
      :::::|||:|::                 ::  :::||||||
22394 CCCGGCCGGTCGGCCCGACGGGGCTGTGTCGCAGGGTGCTGGACATGGTGAAG   22345

  196 ValProGluProAspThrLeuHisAlaGlnIleSerArgLeuGluSerGl   212
      :::|||||||||||||||   :::||:|||||||:|||:|||||||:|||
22344 ACGCCCGAGGGCCGCGGTCTTCTCCGGGCTCGTCGAG...............   22309

  212 yCysThrAlaThrValArgProGly.......................GlyL   222
      |||||||||||||||||||||                           |||
22308 ..............GTCCGTCCGGTCAGCTCGTCCGGGTGGGGCCGGCCG   22272

  222 ysLeuGlnLysArgTyrPheLysProGlnPheProValGlnLysVal   238
      |||:|||||||   :::||::||||||||   ::
22271 GGGTGCCCGGCCAGACGTACTGGCGGCTGGAG................   22240

  239 ValLysGlyLysGluGlnAspLeu.........PheAspArgIleAlaGl   252
      ::::|||||   ::||||||                ::|||||||||||
22239 GCCCGGAGCACACCGACCACCGCCGAACACCCTCCGCCGACCACCTCCGTGA   22190

  252 nValLeuGluAspSerValGluLysHisMetArgAlaAspValThrValG   269
      :::|||||||||:|||   ::|||||||:|||    ::::|||
22189 CCTGCTCGCCGACACCGTCACCCTCAGCTGGTTCTCGGACGTCCCGCTGG   22140

  269 lySerPheLeuSerGlyGlyIleAspSerThrAlaIleAlaProLeuAla   285
      ||:|||||   |||||||||:::|||||      |||||||
22139 GCACGCTCGTCTTCCCGGCGGCCGCCGGCGTCGACTCGTCGGCCGCTGCC   22090
```

```
  286 LysArgHisAsnProAspLeuGluThrPheThrGlyPhe........  299
      |||      ::||        |||       |||
22089 GCCCGCAGCGGCGGCCG...GTGCGCTCCTTCTGCGGTGACTTCCGGGGGC 22043

  300 .........................GluArgGlyGlyTyrSerGluV 307
                               :::|||||||
22042 CGCCGACGCCTTCGCCTCGCACCCGTCCGGCGGGAG......GCGGACG 21999

  307 aIAspValAlaAlaGluAlaAlaAlaIleGlyAlaGluHisIle...  322
      ::       |||  |||  |||        :|||||
21998 CGCCGTACGTGCGGCAGCTGGCCGGGGCACGTCGAGCGGCGACCACCTGGAG 21949

  323 ValLysIleValSerProGluGluTyrAlaAsnAlaIleProLysIleMe  339
      ||| :::  |||||
21948 GTGCTGCTGGACAGCCCCGGC....................  21928

  339 tTrpTyrLeuAspAspProValAlaAspProSerLeuVal............  352
      |||::|:::|||   :::||||||:
21927 .......CTGTCCGACCCGGCGGTCCCGGTGCTGCGCGCCACCG  21885

  353 ...................ProLeuTyrPheVal  357
                          ||||:|:|
21884 ACCTGCCGCCCGGCTACTGGGCGACATGTGGCCCTCGCTCTTCACCTGTTC  21835

  358 AlaAlaGluAlaArgLysHisValValSerGlyValGlyGluGlyValAl  374
      ::|||::::|||  |||  :|||||||||:::|||
21834 TTCCCGGCGTGCGGAGAACGTCACCGTGCGTGCCGGCGGCGAGGCCGC  21785

  374 aAspGluLeuPheGlyGlyTyrThrIleTyrLysGluProLeuSerLeu.  390
      ||||||||||||::::   :::||||||||
21784 GGACGAACTCTTCGGCGGCTACCGGTGTTCCGCAACCCGTTGGCCCCTGC  21735

  391 .....AlaProPheGluLysIleProSerProLeuArgLysLysGlyLeuGly  405
           |||                  |||
21734 GGGCGGCCACCTTC....................  21721

  406 LysLeuSerLysLeuValProAspGlyMetLysGlyLysSerLeuLeuGl  422
                                      |||
21720 ....................................CCGTGGCTCAC  21710

  422 uArgGlySerMetThrMetGluGluArgTyrTyrGlyAsnAlaArgSerP  439
       |||||                     |||||||||:::|||
21709 CCCGGGTTCG.............GGCCGCTACTTCGGCGGCAGCTCGCTGT  21672

  439 heAsnPheGluGlnMetGlnArgValIleProTrpAlaLysArgGluTrp  455
      ||:::                       ::  |||
21671 TCCAGGCGGTCGTGCTGACAAGCTCGACCTGGACGGCGTACCGGGGCGCC  21622

  456 AspHisArgGluValThrAlaProIleTyrArgAlaGlnSerArgAsnPheAs  472
      |||||              :::|||||    :::
21621 CGGTACGCCGAGGCGCTCGCCGAAGTCGCGGTGCTGCGGGTGAGTCGGC  21572

  472 pProValAlaArgMetGln........HisLeuAspLeuPheThrPheTrpM  486
      |||      ||||||             :::||||||||||
21571 GCCCGAGGAGGCGGATCGCGAGATCGGTCACCTGAACCTGACCACCGCTTCG  21522

  486 etArgGlyAspIleLeuValLysAlaAspLysIleAlaAsnMetAlaAsnSer  502
      ::::     |||  |||:::|||:::::::|||||     :::
21521 TGCAG...ACCCTGCTCGACCCGCAAGGACCGGATGAGCAATGGCGGTCGGC  21475

  503 LeuGluLeuArgValProPheLeuAspLysGluValPheLysLysValAlaAlaGl  519
      :::::::                    :::  :::
21474 CTCGAGGTGCGGGTGCCGCTTCTGCGACCACCGGCTCCGTCGACGTACGTCTT  21425

  519 uThrIleProTyrAspLeuLysLysIleAlaAlaAsnGlyThrThrLysTyrAlaL  536
      :::||||||::||||              |||       |||
21424 CAACGTGCCCTGGGAGATGAAGAAGAGTTTCGACGGCCGGGAGAGAGCCTGC  21375

  536 euArgArgAlaLeuGluGlyLnGlnIleIleValProHisHisValLeuHisIleArgLys  552
      ::       |||::||:::|||                 ::|||:||||
21374 TGCGGGCGGCGCTGGAGGGGTTGGCTGCGCTCGGTGCCGCACGGGGTG  21325

  553 LysLeuGlyPheProVal.......................ProMetAr  561
      |||||
```

```
           |||  :::||::::
21324 AAGACCCGTACCCGGCGACCCAGGATCCGGGCTACGAGGGCGCCCTGCG  21275

      561 gHisTrpLeuAlaGly  566
          |  |||||||||
21274 GGCGAACTGGCCGGG  21259
```

seq_name: /SIDS1/gcgdata/geneseqn-embl/NA2001A.DAT:AAF77894

alignment_scores:
```
     Quality:    605.00              Length:      664
     Ratio:        1.622             Gaps:         24
Percent Similarity:  56.175   Percent Identity: 28.313
```

alignment_block:
US-09-786-474-2 x AAF77894   ..

Align seg 1/1 to: AAF77894   from: 1 to: 1872

```
    1 MetCysGlyLeuGlyIleLeuThrAlaAsnGlyAsnAlaAlaGluAlaPh  17
```

```
883  AAGGCGGAGCACGGCGGCGACATCAATTCGTTCTCGGTGGACTTCGTCGG  932
     ||||||||||           :::                    :::
40   ATGTGCGGTCTCGCGGGTTGGTGGATTACACGCGCCAAGCTCGACGACGA  89

300  .GluArgGlyTyrSerGluValAspVal.............         309
     ::: ||    ||||:::
17   evalProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA  34

933  CCAGGCCGAGCAGTTCCGCAGCGACGACCTGCGTCCCGACCAGGACCAGC  982
     |||:||||             ||||||||
90   ATTTCCGGGATCTTCGCCATGACCGATACGCTCCGCTTCGCGCGGCCCGG  139

310  ....AlaAlaGluSerAlaAlaAlaAlaIleGlyValAlaGluHisIleValLys  324
     || ||:|||    |||||||:::||||
983  CGTTCGGCCCTGCTGGCCGCGCCAGTACATCCGGCACGTCATCGGCACCGTG  1032

34   spAspAlaGlyThrTrpHisAspAlaPheGlyPheAsnArg           50
     ::|||| ||||||   :::|||
140  ATGCCAGGGGCATCTGAAGCACCGCAACGCCCTGCTGGTCACCGGCGCG  189

325  IleValSerProGluGlyTyr......................         331
     |||
1033 CTCATCGACAATGCGACGCGGCGGTCTGCGACGAGCGCGGCGAAGAGGTATT  1082

51   LeuSerIleIleAspIleAlaHisSerHisGlnProLeuArgTrpGly..  66
     ||||||:::||||||   ||| |||
190  CTGGCGGTCATCGACCTCAGCGGCGCGGTCAGCGCCGATGTCCTATCGCTT  239

332  .....AlaAsnAlaIleProLysIleMetTrpTyrLeuAspAspProValA  347
     ||| |||:::
1083 CCGGGCCAAGGACGTACCTTTCACCTTCGGCGACATGGAT.........  1122

67   .ProAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyGluIleT  83
     ||| :::         ||||||   :||||
240  TCCCACCGGCCAG......GAGGTCACCCCTGCTCTACACCGGCGAGGTGT  283

347  laAspProSerLeuValProLeuTyrPheValAlaAlaAlaArgLys    363
     :: ||||:::   |||  :::||| |||:::
1123 .........ACCTCGCTGCACCTGATGTTCGGCGGAGAATCCGCCCGG  1158

83   yrAsnTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyrThrPhe  99
     ||| |||:::::||||||||   ||| ||
284  ACAACCACGATGCCCTGCCCGCAGCCGGTTGCGCCCGGCGCCGGACATGAGTTC  333

364  HisValLysValValLeuSerGlyGluGlyAlaAspGlyAlaLeuPheGlyGl  380
     ||| ||   :::||||||||    :::|||   ||||
1159 CATTCCACGGTGGCCATCTCCGGTGAAGGCGCCCACGAGCTGTTCGGTGG  1208

380  yTyrThrIleTyrTyrGluLysGluProLeuSerLeuAlaProPheGluLysIleP  397
     ||||   |||    :::|||   ||||
1209 CTACGGCTGGTTCCGGATCTACCAGGCGGTGGCTGCG...GCCGCGTTCC  1255

397  ro.........SerProLeuArgLysGlyLeuGlyLeuLysLeuSerValLeuLeu  411
     ||  ||| :::|||
1256 CCTGGGCCTCCAGGGTGCGC............................CTG  1278

412  ProAspGlyMetLysGlyLysSerLeuLeuGluLeuGlyArgGlySerMetThrMe  428
     ::: |||  :::||||||
1279 CCGGCCGGC..........TTCATCGACGCCGGTTC............  1305

428  tGluGluArgTyrTyrGlyAsnAlaArgSerPheAsnPheGluGlnMetG  445
     ||||   :::||||||   ||||
1306 .........AACCGCCGCTGCGATCTCCTCCAGTACCAGC  1336

445  lnArgValIleProTrpAlaLysArgGluTrpAspHisArgGluValThr  461
     ||  :: :::|||   |||:::||||||
1337 AGGCCAGCTACGACGATGGGCTGCCCAGGTCGAACACCTGGCCGGCGGAC  1386

462  AlaProIleTyrAlaGlnSerArgAsnPheAspProValAlaArgMetGl  478
     ::: |||||||   :::||||
1387 AGCCCGGAGGAGCGCGGCGGATGCCGGAGTTC.............AG  1418

478  nHisLeuAspLeuPheThrTrpMetArgGlyAspIleLeuValLysAlaA  495
     :::||||   |||||||||||
1419 CCACCTGCATCTGAAGCGCTGGATG...GTGCTGCTGCTCGAACGCAAGG  1465

495  spLysIleAsnMetAlaAsnSerLeuGluLeuArgValProPheLeuAsp  511
     ||:::::   |||||||||||||||:::
1466 ATCGCCTGAGCATGTGCAACGGCCTGGAGGTGCCCTTCCTCACCGAC  1515

512  LysGluValPheLysValAlaGluThrIleProTyrAspLeuLysIleAl  528
     |||  :::||||||   ||| ||
1516 CATGAGCTGGTGGAGTACGTCTACAACGTGCCCCTGGTCGATCAAGAGCCG  1565

528  aAsnGlyThrThrLysTyrAlaLeuArgArgAlaLeuGluGlnIleValP  545
     ::: |||||||   ||||   ||||
1566 GGACGGGCAGGAGAAGTGGCTGCTCAAGCGGGCCTGCGCCGACTATGTCC  1615

545  roProHisValHisArgLysLysGlyLeuPheProValProMetArg    561
     || |||||||||||   :::|||
1616 CGGAGGCCTGCTCAAGCGCCAAGAGCCCTTATCCGACTTCTCCAAC  1665

562  HisTrpLeuAlaGlyAspGluLeuProGlyTrpAlaGlnAspThrIleLy  578
     ||| :::   ||
1666 .....CTCGGCTACGACGGTTTCTCGCGGGGAGCGTCGCGGCCGCCTGCT  1709
```

```
166  sLysThrIleLeu........GluMetAlaGluMetAsnLeuAspL  180
     ::::||||||    :::||||   ||::||
534  CAAGTCCATCCTGCTGGGCGTCCCAATTCGCCAGGCGAAGTTGCGCCGCT  580

166  ProLeuPheThrTyrAlaThrArgThrGluHisSerGlyThrValPheSerSerGluLy  166
     ||||||||    ::: ||||||||||||:::
484  CCGCTGTATTACGCCGCCAGGACGGAGCACTCGGGCACCGTGTTCTCCTC  533

197  ProGlu.....ProAspThrLeuHisAlaGlnIleSerArgLeuGluSe  211
     ||| :::   :::|||
607  TCCCGGGGGCACTTCGCAGAGCACCTGCGAGGTCCAGGAACTGCTCACCCG  656

211  rGlyCysThrHisMetArgAlaAspValThrValGlySerPheLeuS  228
     :: |||||||   |||:::||   ||||
657  CGGGCCACCTGTGTCTGGCGCCAATTCGCACGTTCCGGTGTTCGGCTGCGCT  706

228  yrPheLysAspLysThrIleGlyHisTyrValAspLeuGlnTyrVal  243
     |||   :::|||   |||
707  ACTGGGAG........................GTCGACCTCTGACGCTG  732

244  .........GlnAspLeuPheAspArgIleAlaGlnIleValLeuGluAs  256
     ||| :::   :::|||
733  GCCGACGACCTGCAGAGCACCTGCGAGGCGCACCCCGAACTGGTCACCCG  782

256  pSerValGluLysHisMetArgAlaAspValThrValGlySerPheLeuS  273
     ::::::   |||||||||||:::
783  CGGCCTGGGGCCAATTGCACGCGACGTTCCGGTGTTCGGCTGCTAT  832

273  erGlyGlyIleAspSerThrAlaIleAlaProLeuAlaLeuLysArg....  287
     ||||:::||||||||   |||:::
833  CGGGTGGGGCGATTCGACCCGCCTCGACCGGCGCCCAGCGCATCGCG  882

288  .........HisAsnProAspLeuProLeuThrPheThrThrGlyPhe  299
     :::   ||| ||||||:::   |||
```

alignment_scores:
        Quality: 560.50   Length: 682
        Ratio: 1.519   Gaps: 25
Percent Similarity: 54.106   Percent Identity: 27.713

alignment_block:
US-09-786-474-2 x AAH41225/rev

Align seg 1/1   to reverse of: AAH41225   from: 1   to: 349980

    1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPh 17
      |||||||||||::   ::: ||| |||
209474 ATGTGTGGGATT..........................AATGGGTTTTCTTGGGGTGA 209443

   17 eValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA 34
      ||||||||||||||||||||||||||||||||||||||||||||||||
209442 TGAAGAGTTAGTAAAAAGAATGAACGATGCTATTCGGCATCGTGGACCTG 209393

   34 spAspAlaGlyThrTrpHisAspAlaAlaPheGlyPheAsnArg 50
      |||| |||||||:::::::::::::: ||| |||::|||||||
209392 ATGATGAGGGAGTTTATGTTGATGATAATGTTAGCCTTGGTCACGTTAGG 209343

   51 LeuSerIleAlaSerIleAla..HisSerHisGlnProLeuArgTrpGl 66
      ||::::||||||||||||||| ::||||||||||:::::||||||
209342 CTTGCCATCATTGACCTCTCACCCCAGGGCACCAGCCTATGAAGTAC.. 209295

   66 yProAlaAspGluProAsp......ArgTyrAlaMetThrPheAsnGlyG 81
      ||| ||| || ||::||||||::::|||||||||
209294 .........GAGAAAGATGGTAAAGAGTTTGGATCGTTTACAATGGGG 209255

   81 luIleTyrAsnTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyr 97
      |||||||||||||:::::::::|||||||||||:::: :||||||
209254 AAATTTATAACTTTATGGAAATTCGAAAAGAGCTTGAGGAGAAGGGTTAC 209205

   98 ThrPheAsnThrSerGlyAspGlyGluProIleValGlyPheHisHi 114
      ||||||||| |||::::||:: :::|||||||::||||||||||||::
209204 ACTTTCACCCTCTAACACTGCACCCGAGGTTATCCTGCCGGTCACCTAGA 209155

  114 sTrpGlyGluSerValValGluHisLeuArgGlyMetPheGlyIleAlaI 131
      |||| ||| ||| :::|||||||||||||||||||||||||||
209154 GTGGGGCTTGATTGTGGAGAGTTCAATGGCATGTGGGCTTTCGTGA 209105

  131 leTrpAspThrLysGluLysSerLeuPheLeuAlaArgAspGlnPheGly 147
      ||:::| :::::::||||||||||||||:::::::||||||
209104 TCTATGACAAGAGTAAGAACATCCTGTTTCTAAGCAGGGACAGATTCGGA 209055

  148 IleLysProLeuPheTyrAlaThrThrGluHisGlyThrValPheSerSe 164
      ||| ||||:: :: |||||::::|||||||||||||||:||:||:
209054 ATAAAGCCCCTGTATTATTACTATGACGGCAAAAACATAATCTTCAGCTC 209005

  164 rGluLysThrIleLeuGlu......MetAlaGluMetAsnLeuA 179
      :|||||| :::|| ::::: ||| |::::|||||||
209004 AGAAATCAAGGGAATACTACAACATAACATTCCACGGAAACCGAATGATG 208955

  179 spLeuGlyLeuAsp.......LysArgThrIleGluHisTyrValAsp 192
      ::: |||||||:| ||||||||::|||::||||||
208954 CCGTTATATTTGACTTCCTGTATTACAATCTGCTGACCAC......... 208914

  193 LeuGlnTyrValProGluProAspThrLeuHisAlaGlnIleSerArgLe 209
      |||||||::   ::|||||:::||||||||::||||||
208913 ..........ACGGAGGACACGGTTCTTGAAGGGGATAAAAAGGCT 208879

  209 uGluSerGlyCysThrAlaThrValArgProGly..GlyLysLeuGluG 225
      ||| ||||||||||||||||::|||||||||| ::|||||||||
208878 GATGCCGGGTTATAATGCTGTTTTTGACCTGAACACGAGAAAGTTTGAAG 208829

  225 lnLysArgTyrPheLysProGlnPheProValGlnLysValLysGly 241
      ::::||||||||||||| ::: ||| |||
208828 CTTGGAAGTATTATGACCTAAGA.......GAGCCGAATTGAAAAGCGT 208788

  242 LysGlu................................GlnAspPheAspArgIleAl 251

seq_name: /SIDS1/gcgdata/geneseqn/geneseqn-embl/NA2001A.DAT:AAH41225

seq_documentation_block:
ID   AAH41225 standard; DNA; 349980 BP.
XX
AC   AAH41225;
XX
DT   29-OCT-2001 (first entry)
XX
KW   Pyrococcus abyssi genomic fragment #4.
XX
OS   Hyperthermophilic archaeon; hyperthermophilic protein; ds.
OS   Pyrococcus abyssi.
XX
FH   Key   Location/Qualifiers
FT   misc_feature   1..49980
FT   /*tag= a
FT   /note= "This sequence overlaps with the 3' end of
FT   AAH41224"
FT   misc_feature   300001..349980
FT   /*tag= b
FT   /note= "This sequence overlaps with the 5'end of
FT   AAH41226"
XX
PN   FR2792651-A1.
XX
PD   27-OCT-2000.
XX
PF   21-APR-1999; 99FR-0005034.
XX
PR   21-APR-1999; 99FR-0005034.
XX
PA   (CNRS) CNRS CENT NAT RECH SCI.
PA   (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
PI   Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
PI   Querellou J, Weissenbach J, Saurin W, Heilig R;
XX
DR   WPI; 2001-126236/14.
XX
PT   New nucleotide sequences isolated from Pyrococcus abyssi encode
PT   proteins useful in industry -
XX
PS   Claim 1; Page 429-524; 1657pp; French.
XX
CC   The present invention relates to the genomic sequence of Pyrococcus
CC   abyssi and P. abyssi proteins (see AAB96053-AAB96842). P. abyssi is a
CC   hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal
CC   vents. The present invention provides a fragment of the genomic sequence of P.
CC   abyssi. The 5' end of this sequence overlaps with the 3' end of AAH41224
CC   and the 3' end of this sequence overlaps with the 5' end of AAH41226. The
CC   proteins of the present invention have various potential industrial uses,
CC   since the proteins are stable at very high temperatures, some up to 110
CC   degrees centigrade.
CC   Note: This patent is in the same patent family as WO200065062, which
CC   contains additional sequences as shown in AAB99132-AAB99143,
CC   AAH75903-AAH75920 and AAG66436.
XX
SQ   Sequence 349980 BP; 99421 A; 76616 C; 77444 G; 96499 T; 0 other;

  578 sGluSerGlyThrGluAspIlePheAsn......LysGlnAlaValLeuA 593
      |||:::||||||||| :::|||||| ::: |||
1710 GGAGGACCCGGTCAACCCGGTGTTCGGCATCGTTCCGGAGAGTTCCTGG 1759

  593 spMetLeuAsnGluHisArgAspGly...........ValSerAsp 604
      |||||| :::|||
1760 CCGCCGAACTGGACACTCCGGAGGGGTACTTCAACACCCAGGTGAGCCGC 1809

  605 HisSerArgArgLeuTrpThrValLeuSerPheMetValTrp 618
      ||::: ||| |||:::||||||||:::
1810 CACAAC......CTGGAGACCGGCCACTGGCGCTGGAAGGCTGG 1845

```
208040 AACGCAATGCGGTGGAGCATAGAGAGCAGAGTTCCCTCTGCGACCACGA 207991
               |||||
           ::::|||||||||||   |||
   513 uValPheLysValAlaGluThrIleProTyrAspLeuLysIleAlaAsnG 530
           ::::||:: :::|||  :::||       |||:::  |
207990 GCTCGTTGAATACGTTCTGTCTTCTGCCCCCGAATCCAAGGTTCGGCGG 207941
   530 lyThrThrLysTyrAlaLeuArgAlaLeuArgAlaLeuGlnIleValProPro 546
           || |||||||  |||||||:::||||||:||||:||||:||
207940 GTATTACAAAAGTTCTACTGCGTAAAGGTCTTAAGGGCATCCTGCCGGAT 207891
   547 HisValLeuHisArg...LysLysLeuGlyPheProValProMetArgH1 562
           :::::: |||||     |||||||     ||| :::
207890 GAAATACCGGAATAGGGCTAGCAAGGTAGGGTTTCGACCCCTGACAAAGA 207841
   562 sTrpLeuAlaGlyAspGlu......LeuPheGlyTrpAlaGlnAspThrI 577
           |||                          |||
207840 CATCCTTAAGACTAATGAAGGGCAGAAGTTTGGCTGTGATAGACT 207791
   577 leLysGluSerGlyThrGluAspIlePheAsnLysGlnAlaValLeuAsp 593
           :::::            :::::   |||||||||
207790 CTGAAAGTTCAAAAAACGGCCTTACTGGGATTATCGAAAAAGTATTAAGG 207741
   594 MetLeuAsnGluHisArgAspGluHisSerSerArgArgLeuTr 610
           ||||:: |||||    ||:||:::::::::||||||||
207740 ATGTTGAGGGAGCACGTGTCTGGCAAGAAAAACTGGAGCCAAGAGCTCTG 207691
   610 pThrValLeuSerPheMetValTrpHisGlyIleIlePheValGluAsn 625
           || :::||     :: |||
207690 GAAGGTATAATCATGAGCTGTGGCTCAGGGACTGGATGGATAAT 207645
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA58471

```
seq_documentation_block>
ID  AAA58471 standard; DNA; 58857 BP.
XX
AC  AAA58471;
XX
DT  31-OCT-2000  (first entry)
XX
DE  Nucleotide sequence of the bleomycin (BLM) gene cluster ORFs 8-30.
XX
KW  BLM gene cluster; bleomycin gene cluster; polyketide metabolite;
KW  bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine;
KW  thiazoline; bithiazoline; microbial metabolite; sugar; ss.
XX
OS  Streptomyces verticillus.
XX
FH  Key        Location/Qualifiers
FT  CDS        223..564
FT             /*tag= a
FT             /transl_except= (pos: 1..3, aa: Met)
FT             /note= "ORF 30; encodes AAB07556"
FT  CDS        561..2309
FT             /*tag= b
FT             /transl_except= (pos: 1..3, aa: Met)
FT             /note= "ORF 29; encodes AAB07557"
FT  CDS        2767..3486
FT             /*tag= c
FT             /note= "ORF 28; encodes AAB07558"
FT  CDS        3527..5593
FT             /*tag= d
FT             /transl_except= (pos: 1..3, aa: Met)
FT             /note= "ORF 27; encodes AAB07559"
FT  CDS        5806..12294
FT             /*tag= e
FT             /note= "ORF 26; encodes AAB07560"
FT  CDS        12291..15491
FT             /*tag= f
FT             /note= "ORF 25; encodes AAB07561"
FT  CDS        15488..21013
FT             /*tag= g
FT             /note= "ORF 24; encodes AAB07562"
FT  CDS        21010..24666
```

```
208787           ||||||
      AAGGAAATTTCTGATCCTCAAGGGTCAAGGATCTTTTCCTCAGG..... 208743
          :::||||||||||  |||
   251 aGlnValLeuGluAspSerValGluLysHisMetArgAlaAspValThrV 268
          :::||:: ||||||||||||
208742          GCTGTGAAGATGAGACTTGTTGCCGATGTCCCGG 208709
   268 alGlySerPheLeuSerGlyLylleAspSerThrAlaIleAlaProLeu 284
          ||||||| |||||||||||:::||:||
208708 TGGGTTCTTGCCCACGGTGCCTGCCTTGATACTCCGAGCATAGTC...TGC 208662
   285 AlaLysArgHis......AsnProAspLeuThrPheThrThrGl 298
          |||       :::::||||||||||||
208661 GCCATGAGACACCTCCTCCCGGATGGTGAGATAAAGGTGTTCTCCGCTAGT 208612
   298 yPheGluArgGluGlyTyrSerGluValAspValAlaAlaGluSerAlaA 315
          :::||||       :: :||
208611 TTTTCCGGGCAAG......GAAATTGATGAAAGCAAATATCAG.... 208575
   315 laAlaIleGlyHisGluHisIleLeuValLysIleValSerProGluGlyTyr 331
          ||||||||
208574 ..........................AGAATTGTGTGGAGACGAA........ 208557
   332 AlaAsnAlaIleProLysIleMetTrpTyr................ 341
          |||   ||||||
208556 ..........TGCAGGGTAAGCTGGTACAGGACAACTTTTAACGTCGA 208519
   342                   LeuAspAspProV 346
                            :::||||||||
208518 GGATGTTTTGGCTGATAATCATGATCATAGAACTCAGGAGGACGCCGT 208469
   346 alAlaAspProSerLeuValTyrPheValAlaGluAlaGluAlaArg 362
          :::       ||||  |||     |||||
208468 TTTCGACCCTGAGTATATACCGTCAGTACAGGGTAATGAAGCTCGGCAAGA 208419
   363 LysHis...ValLysValLeuSerGlyGluGlyGluGlyAlaAspGluLeuPh 378
          ::||||:: |||||||||||||||||  :::||
208418 GAAAACGGAATGAAAGTTCTCCGACGGCCAGGAAGGAAGCGATGAGATACT 208369
   378 eGlyGlyTyrTyrThrIleTyr....................... 384
          :::||||||||
208368 TGCCGGATACCACCTACTTCTTCGGCTATTACTTGTCGAGCTCCTTAGGA 208319
   385 ...............LysGluProLeuSerLeuAlaProPheGluLysIle 396
          |||              :::|||
208318 AGTTTAAATGGGCAAACTTTTGAGGGAAATCTTGGCCTACAGCCGGATC 208269
   397 ProSerProLeuArgGlyLeuGlyLeuLysLeuSerLysValLeuProAs 413
          |||:::|||         ::::::
208268 CACGGATCCCTTGTCCCCTCCAAGAACATGGTTCTCTCACAATTCTTCCGGT 208219
   413 pGlyMetLysGlyLysSerLeuLeuGlyArgGlySerMetThrMetGluG 430
          ::: ||| |||:::||||
208218 ATGGGTCACC..AAGAGACTGTGGAGAAGGAGATTCCGTATCTCCGTG 208172
   430 luArgTyrTyrGlyAsnAlaArgSerPheAsnPheGluGlnMetGlnArg 446
          :::||||||||        ::::::
208171 AAGAGTTCATAGAGAAGTTTAAA..........AAACGACCTTCCAAA 208134
   447 ValIleProLeuAlaLysArgGluTrpAspHisArgGluValThrAlaPr 463
          ||||||       :::  |||
208133 GAACTGATATGGAAAATTAAAACTCTCAACCAGGCCTCCTGTTAGCAGA 208084
   463 oIleTyrAlaGlnSerArgAsnPheAsnProValAlaAlaArgMetGlnHisL 480
          ||||      :::  CTACCCTCATC 208061
208083 GACTTACTATCT.........
   480 euAspLeuPheThrTrpMetArgGlyAspIleLeuLysValLysAlaAspLys 496
          ||||||       |||     GAGGACAAA 208041
208060 TCCTGAGGTTT.......
   497 IleAsnMetAlaAsnSerLeuLeuArgValProPheLeuAspLysLysGl 513
          |||  ||
```

```
FT              /*tag= h
FT              /transl_except= (pos: 1..3, aa: Met)
FT   CDS        /note= "ORF 23; encodes AAB07563"
FT              24663..32690
FT              /*tag= i
FT   CDS        /note= "ORF 22; encodes AAB07564"
FT              32893..34830
FT              /*tag= j
FT   CDS        /note= "ORF 21; encodes AAB07565"
FT              34827..35804
FT              /*tag= k
FT              /transl_except= (pos: 1..3, aa: Met)
FT   CDS        /note= "ORF 20; encodes AAB07566"
FT              35818..37302
FT              /*tag= l
FT              /transl_except= (pos: 1..3, aa: Met)
FT   CDS        /note= "ORF 19; encodes AAB07567"
FT              37299..39215
FT              /*tag= m
FT              /transl_except= (pos: 1..3, aa: Met)
FT   CDS        /note= "ORF 18; encodes AAB07568"
FT              39301..47181
FT              /*tag= n
FT   CDS        /note= "ORF 17; encodes AAB07569"
FT              47178..49985
FT              /*tag= o
FT              /transl_except= (pos: 1..3, aa: Met)
FT   CDS        /note= "ORF 16; encodes AAB07570"
FT              49982..51001
FT              /*tag= p
FT              /transl_except= (pos: 1..3, aa: Met)
FT   CDS        /note= "ORF 15; encodes AAB07571"
FT              50998..52386
FT              /*tag= q
FT              /transl_except= (pos: 1..3, aa: Met)
FT   CDS        /note= "ORF 14; encodes AAB07572"
FT              52383..52946
FT              /*tag= r
FT   CDS        /note= "ORF 13; encodes AAB07573"
FT              53018..54190
FT              /*tag= s
FT   CDS        /note= "ORF 12; encodes AAB07574"
FT              54187..55824
FT              /*tag= t
FT   CDS        /note= "ORF 11; encodes AAB07575"
FT              55821..56093
FT              /*tag= u
FT              /transl_except= (pos: 1..3, aa: Met)
FT   CDS        /note= "ORF 10; encodes AAB07576"
FT              56090..57586
FT              /*tag= v
FT              /transl_except= (pos: 1..3, aa: Met)
FT   CDS        /note= "ORF 9; encodes AAB07577"
FT              57583..58857
FT              /*tag= w
FT   CDS        /note= "ORF 8; encodes AAB07578"
XX
PN   WO200040704-A1.
XX
PD   13-JUL-2000.
XX
PF   06-JAN-2000; 2000WO-US00445.
XX
PR   06-JAN-1999; 99US-0115435.
PR   05-FEB-1999; 99US-0118848.
PR   05-JAN-2000; 2000US-0477962.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Shen B,  Du L,  Sanchez C,  Chen M,  Edwards DJ;
XX
PR   WPI; 2000-465974/40.
DR   P-PSDB; AAB07556, AAB07557, AAB07558, AAB07559, AAB07560, AAB07561,
```

```
DR   AAB07562, AAB07563, AAB07564, AAB07565, AAB07566, AAB07567, AAB07568,
DR   AAB07569, AAB07570, AAB07571, AAB07572, AAB07573, AAB07574, AAB07575,
DR   AAB07576, AAB07577, AAB07578.
XX
PT   New bleomycin gene cluster components useful for peptide and/or
PT   polyketide metabolites, especially bleomycin, production and for
PT   chemically modifying biological molecules -
XX
PS   Claim 8; Page 97-135; 162pp; English.
XX
CC   The present sequence represents the BLM (Bleomycin) gene cluster,
CC   containing open reading frames (ORFs) 8-30. The proteins encoded
CC   by the gene cluster are useful for producing peptides and/or polyketide
CC   metabolites, especially bleomycin or bleomycin analogues. They are
CC   also useful for chemically modifying biological molecules to produce
CC   branched methyl groups, and for coupling amino acids and fatty
CC   acids. They may be reacted with an apo-carrier protein and coenzyme A
CC   to produce a holo-carrier protein. The BLM gene cluster or catalytic
CC   domains can be used individually or collectively to produce
CC   thiazolidine, thiazoline, bithiazoline and bithiazoline-containing
CC   microbial metabolites. The BLM gene cluster may also be used to produce
CC   sugars.
XX
SQ   Sequence 58857 BP; 7256 A; 25139 C; 19353 G; 7109 T; 0 other;

alignment_scores:
        Quality: 516.00        Length:     642
          Ratio:   1.487         Gaps:      23
Percent Similarity: 54.050     Percent Identity: 27.726

alignment_block:
US-09-786-474-2 x AAA58471

Align seg 1/1 to: AAA58471   from: 1  to: 58857

       1 MetCysGlyLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPh 17
         :::|||||||:::::|||||||||  :||::::|||||::: ::::::
   37308 GTGTGCGGCATCGTGGCGATCCGTCCGACGCCGACGCTCGACGGC.. 37355

      17 eValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA 34
         ||                ::::|||  :::::||||  ||||||||||
   37356 ....GGTGAACTCACCGCGCGATGGCCGACGCGCCCCGCGGCGGCCCCG 37401

      34 spAspAlaGlyThrTrp......HisAspAlaAspAlaPheGlyPhe 48
         ||  |||||||||        :::||||||:::||:|||:::||||
   37402 ACGGCGAAGGCACCTGGGTCTCGCCCACCGGCCGCCCTGGGCCAC 37451

      49 AsnArgLeuSerIleIleAspIleAlaHisSerHisGlnProLeuArgTr 65
         :::||||||||||||         ::::::|||||||||
   37452 ACCCGGCTCGCCGTGATCGCCCCCGACGCGCCAGCCGGTC...... 37496

      65 pGlyProAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyGluI 82
         ||                              :::||||||
   37497 .......GCCGGCCCGACGCCACCGCCGTCGTCGTCAACGCGCGAGT 37539

      82 leTyrAsnTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyrThr 98
         ::||||:::||      |||::::||   :::|||||
   37540 TCTACGGCTACCGGGAGATCCGCGGAACTGCCGCCGGCTGCCGG 37589

      99 PheAsnThrSerGlyAspGlyGluProIleValValGlyPheHisHisTr 115
         ||::||:::||||||||:::||            :::::::|||
   37590 TTCCGCACCGGCAGCACGCGAG.........ATCGCCCTCCACCTGTA 37630

     115 p.........GlyGluSerValValGluHisLeuArgGlyMetPheGlyI 129
         :                     |||:::||||||||:::|||:::||
   37631 CCTGCGGGACGCGGGCCACTGGAGCCGGTCGAGCCGGCGAGTTCGCCT 37680

     129 leAlaIleTrpAspThrLysLysGluLysSerLeuPheLeuAlaArgAspGln 145
         :::::||||||  :::::  ::::|||  ::|||||  ||||||
   37681 TCGTCCTCTGGGACGAACGCGCCACCCTCTTCGCCCGCCGACCGG 37730

     146 PheGlyIleLysProLeuPheTyrAlaThrThrGluHisGlyThrValPh 162
```

```
37731  ||||||  |||||||||||::|||::                    :::
       TTCGGGCGTCAAACCCCTACTACACCGAGCGCGACACGGGCGGCTCTACGT  37780

162    eSerGluLysLysThrLeuLeuMetAlaGluMetAsnLeuA           179
       :::|||   ::::::||||:||||| |||  ::  ||
37781  CGCCTCGACGGTCAGGGCGTCTCCTGCTCTCCGCCGCCCCGCTGGG      37830

179    spLeuGlyLeuArgpLysArgThrIleGluHisTyrValAspLeuGlnTyr  195
       |||  ||||||
37831  AC.........ACCGCGCCCTTCGCCGCCGCACCTGCAGCTC           37862

196    ValProGluProAsp...ThrLeuHisAlaGlnIleSerArgLeuGluSe   211
       |||||| |||| |||                  :::||
37863  GGCCTGCCCCCGACCGCACCCTCTTCGCCGCATCCGGCAGCTCCCGCC     37912

211    rGlyCysThrAlaThrValArgProGlyLyGlyLeuGlnLeuLysLysArgT  228
       ||||||       |||                    :::
37913  CGGCTGCCACCTCATCGCCGAGCGCCACGGC..ACCCGGTCACCCCCT     37959

228    yrPheLysProGlnPheProVal.....GlnLysValValLysGlyLys    242
       |||||        :::   ||||||||::::::|||:::
37960  ACTGGGACCTCGACTACCCGCCCGGGGAACTCCCGCCCGGGGAAGC       38009

243    GluGlnAspLeuPheAspAsnArgIleAlaLeuGluLeuGluAspSerValGl 259
       ::||||   ||||||       ::::::::::|||:::
38010  CTGGACGACCACCTGGACGCGGTACGCGCCGACGACCGAGCGCCTACG     38059

259    uLysHisMetArgAlaAspValThrValGlySerPheLeuSerGlyGlyI   276
       |||||||||  :::::||   ||||||||||||
38060  GTTGCGTACCGTCGCCGAGCTGCCCCTGCCACCTCAGCCGGCGGCC       38109

276    leAspSerThrAlaIleAlaProLeuAlaLysArgHisAsnProAspLeu   292
       ::||||||:::|||   ||||||| :::||
38110  TGGACTCCTCGCCGCCGCGCCCCGCCACC...CGGCTC              38156

293    LeuThrPheThrThrGlyPhePheValGluArgGluGlyTyrSerGluValAspVa 309
       ::||||||||:::||| |||||:::
38157  ACCGCCTTCACCGTCCGCTTCGACGACCCCTTCGACGAGAGCGCCGT      38206

309    lAlaAlaGluSerAlaAlaAlaAlaIleGlyAlaAlaGluHisHisIleLeuValLysIleV 326
       |||| :::::|||||  :::  |||
38207  CGCCCGGCGACCGACCTGGCCATCGAACAC.                     38243

326    alSerProGluGluTyrAlaAsnAlaIleProLysIleMetTrpTyrTyrLeu 342
       :||||||| ||||||  ::::||
38244  .........CGCGAAGTCGCCTCGGAACGCGCCCACTTCGCCGACCACCTG   38285

343    AspAspProValAlaAspProSerLeuVal.                     352
       :::|||
38286  CGGGACCGTCGCCCGGCGGAGATGGTGCAGGAGAACTCGCACGGCAT      38335

353    .ProLeuTyrPheValAlaAlaGluAlaArgLys...HisValLysValV  368
       ||:||| :::|||::|
38336  CGCCCCGGTACGCCGCACGCACACATCAAGAAGAGGCGGGATTCACCGCCG  38385

368    alLeuSerGlyGlyGlyGlyAlaAspAspProGlyLeuPheGlyGlyTyrThrIleTyr 384
       ||||||:::||||||||| |||||
38386  TCCTCGCCCGGGAGGGCGGCGGGGACGAACTGTTCCTCGGCTACCCCCAGTTC 38435

385    LysGluProLeuSerGluAla.                              391
       :::::::::||||:::||
38436  CGCAAGGACCTGACGGTCCAGCCTGTCCGCGACGCCCGGACAAGGCCGA    38485

392    .........ProPheGluLeuProSerProLeuA                  401
       ::||||  ||||
38486  CCGGGCTACGCCCGGCTGTCGCGCCGGCGCTCCTGCCGCCGTACCTGC     38535

401    rgLysGlyLeuGlyLeuSerLeuLysSerValLeuProAlaAspGlyMetLysGly 417
       |||||| :::
38536  GCACCCTCCTCGGCACCCGGC...TTCCTGCCCTCCTCGGATCGTCGAC    38582

418    LysSerLeuLeuGluArgGlySerMetThrMetGluGluArgTyrTyrGl   434
       :::   |||
```

```
38583  CGGCACCTGGCGTCACCAGCGCCGTCCGCCGCCTGCTCCGC.........  38624

434    yAsnAlaAargSerPheAsnPheGluGlnMetGlnArgValIleProTrpA  451
                                                         |||
38625  .................................................CCCGACT  38631

451    lALysArgGluTrpAspHisArgGluValThrAlaProIleTyrAlaGln   467
       :::   ::::::|||| |||||:::  ||
38632  TCGCCGCCGAACTGGCCCGCCGAGCGCCCCTGCTCGCGCCC           38681

468    SerArgAsnPhe.........AspProValAlaArgMetGlnHisIle     480
       ||| :::
38682  GGGCCGGCCGCTGCTCGCCGGCGCCGGCGCCGCCACCAGCCACCTACCT    38731

480    uAspLeuPheThrTrpMetArgGlyAspIleLeuArgLeuValLysAlaAspLysSI 497
       |    ||||||:::|||  |||::::::
38732  CTTCGCCAAGACCTGGTCGCCGGC..TACCTGCTCCGCCGGAACGCC      38778

497    leAsnMetAlaAsnSerLeuGluLeuArgValProPheLeuAspLysGlu   513
       :::  ||  ::|||::::::|||:::|||   ::|:::
38779  TCGACGGCGCCAGGCGCGTGAGGTGCGGGCGTGCGCCCCTTCGACCACCAC  38828

514    ValPheLysValAlaGluThrIlePro.........TyrAspLeuLysIl  527
       :::|||         |||
38829  CTCTTCGACCTCGTCCGGCACACCCGCCGGCTGGTACGAC.........   38870

527    eAlaAsnGlyThrThrLysTyrAlaAlaLeuArgArgAlaAlaLeuGluGlnIleV 544
       ::  |||||||  |||||  ||
38871  .AAGGACGGCACCGGCAAGTACCCGCTGCGGGCCGCCATGGCCACCGGC    38919

544    alProProHisValLeuHisArgLysLeuLysGlyPheProValProMet  560
       :||| :::||| ::: ||||:::|||
38920  TGCCGGCGGAGGTGACCAGGGCGCCAAACAGGGCTTCCTGGCACCTCCG    38969

561    ArgHisTrpLeuAlaGlyAspGluLeuProGlyTrpProAlaGlnAspThrIl 577
       :::::||||    |||
38970  .......ATGGCCGACGACGACACCCTC.........CTCCGACGCCCT    39001

577    eLysGlu.........SerGlyThrGluAspIlePheAsnL          588
       :::|||   :::::||  :::||
39002  GCGGGAACGCGCTCGCCGGACCGGCGCGGCGGCGACGACCCCTTCTTCGACC 39051

588    ysGlnAlaValLeuAspMetLeuAsn                          596
       :::|||||| :::|||
39052  CGCACCGGGTCCCGCCCCTGCTCGAC                          39077
```

PR	25-MAR-1999;	99US-0126264.	PR	16-JUL-1999;	99US-0144086.
PR	29-MAR-1999;	99US-0126785.	PR	19-JUL-1999;	99US-0144325.
PR	01-APR-1999;	99US-0127462.	PR	19-JUL-1999;	99US-0144331.
PR	06-APR-1999;	99US-0128234.	PR	19-JUL-1999;	99US-0144332.
PR	08-APR-1999;	99US-0128714.	PR	19-JUL-1999;	99US-0144333.
PR	16-APR-1999;	99US-0129845.	PR	19-JUL-1999;	99US-0144334.
PR	19-APR-1999;	99US-0130077.	PR	19-JUL-1999;	99US-0144335.
PR	21-APR-1999;	99US-0130449.	PR	20-JUL-1999;	99US-0144352.
PR	23-APR-1999;	99US-0130510.	PR	20-JUL-1999;	99US-0144632.
PR	23-APR-1999;	99US-0130891.	PR	20-JUL-1999;	99US-0144884.
PR	28-APR-1999;	99US-0131449.	PR	21-JUL-1999;	99US-0144814.
PR	30-APR-1999;	99US-0132048.	PR	21-JUL-1999;	99US-0145086.
PR	30-APR-1999;	99US-0132407.	PR	21-JUL-1999;	99US-0145088.
PR	04-MAY-1999;	99US-0132484.	PR	22-JUL-1999;	99US-0145085.
PR	05-MAY-1999;	99US-0132485.	PR	22-JUL-1999;	99US-0145087.
PR	06-MAY-1999;	99US-0132486.	PR	22-JUL-1999;	99US-0145089.
PR	07-MAY-1999;	99US-0132487.	PR	22-JUL-1999;	99US-0145192.
PR	11-MAY-1999;	99US-0132863.	PR	23-JUL-1999;	99US-0145145.
PR	14-MAY-1999;	99US-0134256.	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134218.	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	99US-0134219.	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	99US-0134221.	PR	27-JUL-1999;	99US-0145913.
PR	18-MAY-1999;	99US-0134370.	PR	27-JUL-1999;	99US-0145918.
PR	18-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145919.
PR	19-MAY-1999;	99US-0134941.	PR	28-JUL-1999;	99US-0145951.
PR	20-MAY-1999;	99US-0135124.	PR	02-AUG-1999;	99US-0146386.
PR	21-MAY-1999;	99US-0135353.	PR	02-AUG-1999;	99US-0146388.
PR	24-MAY-1999;	99US-0135629.	PR	03-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	99US-0136021.	PR	03-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	99US-0136392.	PR	04-AUG-1999;	99US-0147204.
PR	28-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	99US-0137222.	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	18-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	21-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	22-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	16-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.

us-09-786-474-2.p2n.rng

PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

alignment_scores:
        Quality: 468.00   Length: 617
        Ratio: 1.436   Gaps: 27
Percent Similarity: 52.836   Percent Identity: 26.580

alignment_block:
US-09-786-474-2 x AAC40911   .

Align seg 1/1 to: AAC40911 from: 1 to: 2164

   1 MetCysGlyLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAla..   16
     ||||||||:::||::||:   :::::::|||
 196 ATGTGTGGAATACTTGCCGTGTTAGGATGTTCCGATGATTCTCAGGCCAA 245

  17 ......PheValProAlaLeuGluArgAlaLeuProCysMetArgHisA   31
            ||| |||:::||| ||| ||||||
 246 GAGAGTTCGTGTTCTCGAGCTTCTCCGAGATTG........AGGCACA   286

  31 rgGlyProAspAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGly   47
     ||||||||| :::||| |||||::||||||||||||||
 287 GAGGACCTGACTGGAGTGGATTATATCAGAACGGAGATAATTACTTCGGCC 336

  48 PheAsnArgLeuSerIleIleAspIleAlaHisSerHisGlnProLeuAr   64
     :::|||||||:::|||||| ||| |||||||:::
 337 CATCAACGCTCTTGCCGTCATCGATCCTGCTTCCGGTGATCAACCTCTT.. 384

  64 gTrpGlyProAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyG   81
                 ::||| :::::||| |||||||
 385 ........TTCAACGGAGGACAAGACCATTGTTGTCACGGTGAACGGAG   424

  81 luIleTyrAsnTyrValGluLeuArgLysGluSerAspLeuGlyTyr   97
     |||:::|||:::||||||||||||||||| :::
 425 AGATTTATAACCATGAGGAGCTGAGAAAACGTCTGAAGAAT......CAC 468

  98 ThrPheAsnThrSerGlyAspGlyGluProIleValValGlyPheHisHi   114
     ||||||::|||||||||||||||||||| ||| :||
 469 AAGTTCCGTACTGGGAGTGATGTGAAGTCAATTGCTCACTTGTACCAGGA 518

 114 sTrpGlySerValValGluHisLeuArgGlyMetPheGlyIleAlaI   131
     :::||| ::: ||||||||| ||| ||||:::|||:::||
 519 GTATGGTGTGGAGTTTGTTGATATGTTGGAATGGGAATCTTCTCCTTTCTGT 568

```
1275 .................................................... 1275

 431 ArgTyrTyrGlyAsnAlaArgSerPheAsnPheGluGlnMetGlnArgVa  447

1275 .................................................... 1275

 447 IleProTrpAlaLysArgGluTrpAspHisArgGlyValThrAlaProI   464
                    |||:::||:::|        ||:::|||::|
1276 ..........AAGAAAGAGTTT...CACCAGAAACTTGTCGCAAGA      1309

 464 leTyrAlaGlnSerArgAsnPheAspProValAlaArgMetGlnHisLeu  480
     ||   |||   :::      .    :: :::
1310 TCAAGGCTCTTCACAAG...TATGAC.........                 1332

 481 AspLeuPheThrTrpMetArgGlyAspIleLeuValLysAlaAspLysIl  497
                                     :::::|||:::::|||
1333 ................TGTCTAAGAGCCAACAAATC                1352

 497 easnMetAlaAsnSerLeuGluLeuArgValProPheLeuAspLysGluV  514
          ::|||    ||||||||  |||||||||||||||||
1353 TACCTCTGCCTTGGACTAGAGGCACGTGTTCCTTTCCTTGACAAAGACT   1402

 514 alPheLysValAlaGluThrIleProTyrAspLeuLysIle.........  527
          ||:::|  |||    ||  :::::|  ||||
1403 TCATCAACACAGCTATGTCTCTCGACCCTGAATCCAAGATGATCAAGCCA  1452

 528 ...AlaAsnGlyThrThrLysTyrAlaLeuArgArgAlaLeuGluGln..  542
          :::        ||:::|||||||||||||||||
1453 GAGGAAGGAAGGATCGAGAAATGGGTTCTAAGGAGAGCCTTTGACGACGA  1502

 543 .........IleValProHisValLeuHisLysArgLysLysLeuGlyP   556
                    |||  ||:::||:::::|||::::||
1503 AGAACGTCCTTATCTACCAAAACACATTCTCTACAGACAGAAAGAACAGT  1552

 556 heProValProMetArgHisTrpLeuAlaGlyAspGluLeu...PheGly  571
          :::|        :::  ::::::|||::|
1553 TC..........AGTGATGGTGTTGGCTACAGT                   1575

 572 TrpAlaGlnAspThrIleLysGluSerGlyThrGluAspIlePheAsnLy  588
     |||   |||  :::|||   :::|||||||  :::||
1576 TGCATC...GATGGCCTGAAAGATCVCGCTGCTCAAAATGTCAATGACAA  1622

 588 s   588
     |
1623 G   1623
```

seq_name: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2000.DAT:AAC50959

us-09-786-474-2.p2n.rng

Tue Sep 24 09:40:56 2002

```
PR   16-JUL-1999;  99US-0144085.        PR   07-OCT-1999;  99US-0158029.
PR   16-JUL-1999;  99US-0144086.        PR   08-OCT-1999;  99US-0158232.
PR   19-JUL-1999;  99US-0144325.        PR   12-OCT-1999;  99US-0158369.
PR   19-JUL-1999;  99US-0144331.        PR   13-OCT-1999;  99US-0159293.
PR   19-JUL-1999;  99US-0144332.        PR   13-OCT-1999;  99US-0159294.
PR   19-JUL-1999;  99US-0144333.        PR   13-OCT-1999;  99US-0159295.
PR   19-JUL-1999;  99US-0144334.        PR   14-OCT-1999;  99US-0159329.
PR   20-JUL-1999;  99US-0144335.        PR   14-OCT-1999;  99US-0159330.
PR   20-JUL-1999;  99US-0144352.        PR   14-OCT-1999;  99US-0159331.
PR   20-JUL-1999;  99US-0144632.        PR   14-OCT-1999;  99US-0159637.
PR   20-JUL-1999;  99US-0144884.        PR   14-OCT-1999;  99US-0159638.
PR   21-JUL-1999;  99US-0144814.        PR   18-OCT-1999;  99US-0159584.
PR   21-JUL-1999;  99US-0145086.        PR   21-OCT-1999;  99US-0160741.
PR   21-JUL-1999;  99US-0145088.        PR   21-OCT-1999;  99US-0160767.
PR   22-JUL-1999;  99US-0145085.        PR   21-OCT-1999;  99US-0160768.
PR   22-JUL-1999;  99US-0145087.        PR   21-OCT-1999;  99US-0160770.
PR   22-JUL-1999;  99US-0145089.        PR   21-OCT-1999;  99US-0160814.
PR   22-JUL-1999;  99US-0145192.        PR   21-OCT-1999;  99US-0160815.
PR   23-JUL-1999;  99US-0145145.        PR   22-OCT-1999;  99US-0160980.
PR   23-JUL-1999;  99US-0145218.        PR   22-OCT-1999;  99US-0160981.
PR   23-JUL-1999;  99US-0145224.        PR   22-OCT-1999;  99US-0160989.
PR   26-JUL-1999;  99US-0145276.        PR   25-OCT-1999;  99US-0161404.
PR   27-JUL-1999;  99US-0145913.        PR   25-OCT-1999;  99US-0161405.
PR   27-JUL-1999;  99US-0145918.        PR   25-OCT-1999;  99US-0161406.
PR   27-JUL-1999;  99US-0145919.        PR   26-OCT-1999;  99US-0161359.
PR   28-JUL-1999;  99US-0145951.        PR   26-OCT-1999;  99US-0161360.
PR   02-AUG-1999;  99US-0146386.        PR   26-OCT-1999;  99US-0161361.
PR   02-AUG-1999;  99US-0146388.        PR   28-OCT-1999;  99US-0161920.
PR   02-AUG-1999;  99US-0146389.        PR   28-OCT-1999;  99US-0161992.
PR   03-AUG-1999;  99US-0147038.        PR   28-OCT-1999;  99US-0161993.
PR   04-AUG-1999;  99US-0147204.        PR   29-OCT-1999;  99US-0162142.
PR   04-AUG-1999;  99US-0147302.
PR   05-AUG-1999;  99US-0147192.
PR   05-AUG-1999;  99US-0147260.
PR   06-AUG-1999;  99US-0147303.
PR   06-AUG-1999;  99US-0147416.
PR   09-AUG-1999;  99US-0147493.
PR   09-AUG-1999;  99US-0147935.
PR   10-AUG-1999;  99US-0148171.
PR   11-AUG-1999;  99US-0148819.
PR   12-AUG-1999;  99US-0148341.
PR   13-AUG-1999;  99US-0148565.
PR   13-AUG-1999;  99US-0148684.
PR   16-AUG-1999;  99US-0149368.
PR   17-AUG-1999;  99US-0149175.
PR   18-AUG-1999;  99US-0149426.
PR   20-AUG-1999;  99US-0149722.
PR   20-AUG-1999;  99US-0149723.
PR   20-AUG-1999;  99US-0149929.
PR   23-AUG-1999;  99US-0149902.
PR   25-AUG-1999;  99US-0149930.
PR   26-AUG-1999;  99US-0150566.
PR   27-AUG-1999;  99US-0150084.
PR   27-AUG-1999;  99US-0151065.
PR   30-AUG-1999;  99US-0151066.
PR   31-AUG-1999;  99US-0151080.
PR   01-SEP-1999;  99US-0151303.
PR   07-SEP-1999;  99US-0151438.
PR   10-SEP-1999;  99US-0151930.
PR   13-SEP-1999;  99US-0152363.
PR   15-SEP-1999;  99US-0153070.
PR   16-SEP-1999;  99US-0153758.
PR   20-SEP-1999;  99US-0154018.
PR   23-SEP-1999;  99US-0154039.
PR   24-SEP-1999;  99US-0154779.
PR   28-SEP-1999;  99US-0155139.
PR   29-SEP-1999;  99US-0155486.
PR   04-OCT-1999;  99US-0155659.
PR   05-OCT-1999;  99US-0156458.
PR   06-OCT-1999;  99US-0156596.
PR                 99US-0157117.
PR                 99US-0157753.
PR                 99US-0157865.
```

```
alignment_scores:
        Quality:  467.00                          Length:    617
        Ratio:    1.437                           Gaps:       27
Percent Similarity: 52.674       Percent Identity: 26.580

alignment_block:
US-09-786-474-2 x AAC50959

Align seg 1/1    to: AAC50959      from: 1    to: 2162

  1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAla..  16
    !!!!!!!!!!!!::!!!    :::::::  :!!!!!!!!!  :::::!!!
195 ATGTGTGGAATACTTGCCCTGTTAGGATGTTCCATGATTCTCAGGCCAA  244

 17 .......PheValProAlaLeuGluArgAlaLeuProCysMetArgHisA  31
            !!!   !!!::::!!!   !!!  !!!!!!!
245 GAGAGTTCCTGTCTTCTTGAGCTTTCTCCGCAGATTG........AGGCACA  285

 31 rgGlyProAspAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGly  47
    !!!!!!!!!  ::::!!!  :::::::::!!!   :::::!!!
286 GAGGACCTGACTCGGATGGCTTATATCAGAACCGAGATAATTACTTGGCC  335

 48 PheAsnArgLeuSerIleIleAspIleAlaHisSerHisGlnProLeuAr  64
    ::!!!!!!!!!!::::!!!!!!!  !!!  :::  !!!!!!!!!
336 CATCAACGTCTTGCCGTCATCGATCCTGCTTCCGTGATCAACCTCTT..  383

 64 gTrpGlyProAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyG  81
    :::!!!  ::::::::!!!!!!!!!!  ::::!!!  !!!!!!!!!
384 ........TTCAACGAGGACAAGACCATTGTTGTGCACGGTGAACGGAG  423

 81 luIleTyrAsnTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyr  97
    !!!!!!!!!:!  ::::::::!!!::!!!::::::     :::
424 AGATTTATAACCATGAGGAGCTGAGAAAACGTCTGAAGAAT......CAC  467

 98 ThrPheAsnThrSerGlyAspGlyGluProIleValValGlyPheHisHi  114
    !!!::!!!!:!:::!!!  !!!  !!!::!!!
468 AAGTTCCGTACTGTAGTCATTGTGAAGTCATTGCTCACTTGTACGAGGA  517

114 sTrpGlyGluSerValValGluHisLeuArgGlyMetPheGlyIleAlaI  131
    :::!!!  !!!  !!!::::!!!!!!!!!::!!!  !!!  !!!
518 GTATGGTGTGGGATTTTGTTGATATGTTGGATGGAATATTCCTTTGTGT  567
```

```
414  lyMetLysGlyLysSerLeuLeuGluArgGlySerMetThrMetGluGlu 430
                                                                                    1274  .................................................. 1274
431  ArgTyrTyrGlyAsnAlaAlaArgSerPheAsnPheGluGlnMetGlnArgVa 447
                                                                                    1274  .................................................. 1274
447  lIleProTrpAlaLysArgGluTrpAspHisArgGluValThrAlaProI 464
          ||||||           ||:::||:|||||                  ||                       1275  .........AAGAAAGAGTTT...CACCAAGAAACTTGTCGCAAGA 1308
464  leTyrAlaGlnSerArgAsnPheAspProValAlaArgMetGlnHisLeu 480
          ||  |||                         :::  :::|||                               1309  TCAAGGCTCTTCACAAG...TATGAC                 1331
481  AspLeuPheThrTrpMetArgGlyAspIleLeuValLysAlaAspLysIl 497
                                                                     ||||:          1332  .....................................TGTCAAGAGCCAACAAATC 1351
497  eAsnMetAlaAsnSerLeuGluLeuArgValProPheLeuAspLysGluV 514
          :::   |||                 |||||||||||||||||||||||                          1352  TACCTCTGCCTTGGACTAGAGGCACGTGTTCCTTCCTTGACAAAGACT 1401
514  alPheLysValAlaGluThrIleProTyrAspLeuLysIle.......... 527
          :::||||:|||:|||                            |||:|||                        1402  TCATCAACACAGCTATGTCTCTCGACCCTGAATCCAAGATGATCAAGCCA 1451
528  ...AlaAsnGlyThrThrLysTyrAlaLeuArgArgAlaLeuGluGln.. 542
          |||||||||||||||||||||||||||||||||||||||||                                 1452  GAGGAGGAAGGATCGAGAAATGGGTTCTAAGGAGGAGCCTTTGACGACGA 1501
543  .......IleValProProHisValLeuHisArgLysLysLeuGlyP 556
          ::| |||      |||||||::|||||||||||||||||||                                  1502  AGAACGTCCTTATCTACCAAAAACACATTCTACAGACAACAAAAGAACAGT 1551
556  heProValProMetArgHisTrpLeuAlaGlyAspGluLeu...PheGly 571
                        :|||||     :::                  ::|:::|                     1552  TC..................................AGTGATGGTGTTGGCTACAGT 1574
572  TrpAlaGlnAspThrIleLysGluSerGlyThrGluAspIlePheAsnLy 588
          ||      |||:::||||                            :::||||                      1575  TGGATC...GATGGCCTGAAAGATCACGCTGCTCAAAATGTCAATGACAA 1621
588  s 588
          ||
1622  G 1622
```

seq_name: /SIDS1/gcgdata/geneseqn-embl/NA1999.DAT:AAV81946

seq_documentation_block:
ID   AAV81946 standard; DNA; 40138 BP.
XX
AC   AAV81946;
XX
DT   19-OCT-1999 (first entry)
XX
DE   V. marinus PKS-like cluster comprising ORFs 6,7,8 and 9.
XX
KW   Polyketide-like synthesis; PKS; PKS-like gene; PUFA; DHA; transgenic;
KW   poly-unsaturated fatty acid; eicosapentenoic acid; docosahexanoic acid;
KW   EPA; oil; dietary supplement; infant feeeding formulation; malnutrition;
KW   intravenous feeding formulation; cooking oil; fat; anti-inflammatory;
KW   cholesterol; open reading frame; ORF; ss.
XX
OS   Vibrio marinus.
XX
FH   Key   Location/Qualifiers
FT   CDS   17394..25352
FT   /*tag= a
FT   /product= "ORF 6"
FT   CDS   25509..28160
FT   /*tag= b

```
131  leTrpAspThrLysGluSerLeuPheLeuAlaArgAspGlnPheGly 147
          :::||||||||||||||           :::||||||||||||||    :::||                      568  TGCTCGACACAGAGATAACTCTTTCATGGTGGCTCCGTGATGCGATTGGT 617
148  IleLysProLeuPheTyrAlaThrThrGlyHisGlyThrValPhe...Se 163
          :::         |||                |||||||||||||                                618  GTCACTTCGCTCTACAATGGTTGGGACTAGACGGATCTCTGTGGGATATC 667
163  rSerGluLysLysThrIleLeuGluMetAlaGluGluMetAsnLeuAspL 180
          ||||||  |||                                                                 668  TTCAGAGATGAAA...                             680
180  euGlyLeuAspLysProArgThrIleGluHisTyrValAspLeuGlnTyrVal 196
          ||::||:|||                |||||                                             681  .GGCCTAAACGAT...GACTGTGAGCATTTC...         707
197  ProGluProAspThrLeu........HisAlaGlnIleSerArgLeuGl 210
          :::|||                      |||     |||                                     708  .......GAAACGTTCCTCCAGGTCATTTTATTCAAGCAAATTA.. 746
210  uSerGlyCysThrAlaThrValArgProGlyGlyLysLeuGlyGlnLysA 227
          |||||                                   ||||                                747  .........GGAGGG.........TTTAAGC               759
227  rgTyrPheLysProGlnPheProValGlnIleLysValValLysGlyLysGlu 243
          :::||||:||:|||||             ::  :::                                        760  AATGGTATAATCCTCCTGGTTCAATGAATCTGTTCCTTCAACGCCCTTAT 809
244  GlnAspPheAspAsnArgIleAlaGlnLeuValLeuGluAspSerValGluLy 260
          :::|||       |||                    :::|||                                  810  GAGCCTCTT......GCGATAAGACACGCGCCTTTGAAAACGTGTGATTAA 853
260  sHisMetArgAlaAspValThrValGlySerPheLeuSerGlyLysGlyIleA 277
          ||::||||| ::||||||   |||     :::|||||||||||||||:::||                        854  GCGGTTGATGACTGATGTTCCATTGGAGTTTTGCTCTCTGGTGGTCTTG 903
277  spSerThrAlaIleAlaProLeuAlaLysArgHis.............. 288
          |||||||                  ::: :::||||||                                      904  ATTCTTCCCTTGTTGCCTCCATCACTGCACGTCACTTGGCCGGTACTAAG 953
289  ..........AsnProAspLeuThrProPheThrThrGlyPheGl 300
          :::|||:||||          ::|||                                                  954  GCGGCTAAGCAATGGGGTCCTCCAGCTCCATTCCTTTTGCGTTGGTCTT.. 1001
300  uArgGluGlyThrTyrSerGluValAspValAlaAlaGluSerAlaAlaAlaI 317
          ||||||          :::  ::: :::||||                                            1002  ....GAGGGCTCACCGGACTTGAAGGCAGGAAAGAGGTGGCGGAATATT 1047
317  leGlyAlaGluHisIleValLysIleValSerProGluGluTyrAlaAsn 333
          ::||||| |||                    :::                                          1048  TGGGACGGTGCACCACGAGTTCCACTTCTCGGTGCAGGACGGGATTGAT 1097
334  AlaIleProLysIleMetTrpTyrLeuAsp.....AspProValAlaAs 348
          ||||||                            |||                                       1098  GCCATAGAGGATGTGATTTACCATGTTGAGACCTATGTGACGACTAT 1147
348  pProSerLeuValProLeuProLeuSerLeuAspValAlaAlaAlaArgLys...H 364
          :::     :::||||                     :::  :::                                1148  CAGAGCGACGCACCAGATGTTCTTGATGTCCCGGAAAATCAAGTCTCTAG 1197
364  isValLysValLeuSerGlyGluGlyAlaAlaAspGluLeuPheGlyGly 380
          ||||||||||||||||||||||||||||||||||                                          1198  GGGTCAAGATGGTTCTCTCCGGCAAGGTGCGGACGAGATCTTTGGAGGG 1247
381  TyrThrIleTyrLysGluProLeuProPheGluLysIlePror 397
          |||     :::|||                        :::||                                 1248  TAC...CTCTAT.............TTCCACAAGGCACC 1270
397  oSerProLeuArgLysGlyLeuGlyLysLeuSerLysValLeuProAspPG 414
          :::                                                                        1271  TAAC.............................          1274
```

<div style="display:none">
147
617
163
667
180
680
196
707
210
746
227
759
243
809
260
853
277
903
288
953
300
1001
317
1047
333
1097
348
1147
364
1197
380
1247
397
1270
414
1274
</div>

```
FT   CDS              /product= "ORF 7"
FT                    28209..34265
FT                    /tag= c
FT   CDS              /product= "ORF 8"
FT                    34454..36118
FT                    /tag= d
FT                    /product= "ORF 9"
XX
PN   WO9855625-A1.
XX
PD   10-DEC-1998.
XX
PF   04-JUN-1998; 98WO-US11639.
XX
PR   04-JUN-1997; 97US-0048650.
XX
PA   (CALJ ) CALGENE LLC.
XX
PI   Facciotti D, Lassner M, Metz JG;
XX
DR   WPI; 1999-070271/06.
XX
PT   New nucleic acid encoding polyketide-like synthesis enzymes of
PT   Vibrio marinus - and transformed plants and microbes that produce
PT   polyunsaturated fatty acids, useful as pharmaceuticals and food
PT   supplements
XX
PS   Example 1; Fig 5; 153pp; English.
XX
CC   The invention provides polyketide-like synthesis (PKS)-like genes that
CC   are used for the production of long chain poly-unsaturated fatty acid
CC   (PUFA) productions. Genes responsible for eicosapentenoic acid (EPA)
CC   production in Shewanella putrefaciens and novel genes associated with the
CC   production of docosahexanoic acid (DHA) in Vibrio marinus are used to
CC   generate transgenic plants that can express transgenes encoding PKS-like
CC   genes associated with PUFA production. The PKS-like genes are used to
CC   transform plants and microbial cells to give recombinants having altered
CC   contents of PUFA (specifically DHA and EPA). Oils from these plants are
CC   useful as dietary supplements (in infant feeding formulations, to give a
CC   PUFA profile closer to that of human milk; for treating malnutrition; in
CC   intravenous feeding formulations; in cooking oils, fats etc.), also as
CC   anti-inflammatory agents and for reducing cholesterol levels. Fragments
CC   from the genes are useful as probes to isolate related molecules or to
CC   detect organisms that express PKS-like genes. The method facilitates
CC   large scale production of PUFA by providing new pathways for their
CC   synthesis or suppressing interfering pathways. Expression of PUFA in
CC   seeds allows simple recovery, as oil which can be engineered to have a
CC   particular PUFA profile. Expression in microbes also allows simple
CC   recovery and control of PUFA profile and is not subject to external
CC   variables such as weather or food supply. The present sequence represents
CC   an approximately 40 kb PKS-like cluster DNA fragment from V. marinus
CC   comprising open reading frames (ORFs) 6, 7, 8 and 9.
XX
SQ   Sequence 40138 BP; 11845 A; 7635 C; 8758 G; 11900 T; 0 other;
```

```
alignment_scores:
      Quality:  463.50          Length:  589
      Ratio:    1.525           Gaps:    20
Percent Similarity: 51.613   Percent Identity: 25.127

alignment_block:
US-09-786-474-2 x AAV81946/rev

Align seg 1/1 to reverse of: AAV81946  from: 1  to: 40138

    1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPh 17
      ||||||||||:::||:||||||    :::::|||       ::
11291 ATGTGCTCAATATTCGGAATTCTAGACATCAAATCAGACATTAAA..... 11247

   17 eValProAlaLeuGluArgAlaLeu.......ProCysMetArgHisA 31
      ||||||||||||:|||||||||||       ||:||||||
11246 ...CCTCTTCCGGACAACAAGCACTAGAACTATCAAAGTTACTACGTCACC 11201
```

```
 31 rgGlyProAspAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGly 47
    ||||||||||::::||       :::|||       :::|||||||||
11200 GCGGCCCAGATTGGTCAGTATCTATACAAGTGACAATGCTATTTTAGTT 11151

 48 PheAsnArgLeuSerIleIleAspIleAlaHisSerHisGlnProLeuAr 64
    :::||||||||:::|||:::|| :::::||| ||||:||       |||||||
11150 CATGAACGTCTTGCTATTGTCGATGTAAATAATGGTGCACAGCCGCTA.. 11103

 64 gTrpGlyProAlaAspGluProAspArgTyrAlaMetThrProAsnGlyG 81
    ::|||       :::::::|||       |||||||    :::::|||||||
11102 ...TATAATGAAGAAAGACACCACGTACTTGCTGTCGTTAATGGTG 11063

 81 luIleTyrAsnTyrValGlyGluLeuArgLysGluSerAspLeuGlyTyr 97
    |||:::|||    ::::|||||| |||       :::::      |::  :::
11062 AAATTTATAACCACAAAGATTTAAAGAAAACGCTG... AACGTAGACTTT 11016

 98 ThrPheAsnThrSerGlyAspGlyAspProIleValValGlyPheHisH1 114
    |||:::|||:::||:::||| |||       :|| |||       :::|||
11015 GAATTCCAAACAGAATCAGATTCGCAAGTGATACTTGGCCTATATAAAGA 10966

114 sTrpGlyGluserValValGluHisLeuArgGlyMetPheGlyIleAlaI 131
    : ||       ::::::   |||:::||::::::|||:::::|||:::|||
10965 AAAAGGCACACAATTTTAGATGATCTAAATGGTATTTTGCTTTCGCTT 10916

131 leTrpAspThrLysGluLysSerLeuPheLeuAlaArgAspGlnPheGly 147
    ||       :::|||       ::::::|||||| |||       :::|||
10915 TATATGACGAAACAGAGGACGCCTACCTGATTGCCGGTGACCATATCGGT 10866

148 IleLysProLeuPheTyrAlaThrThrGluHisGlyThrValPhe...Se 163
    |||     ::::::      |||||:::      :::|||||||      :::
10865 ATCATCCCTCTTTATACGGGCATGATGAACACGGTAACTTTTATGTTGC 10816

163 rSerGluLysThrIleLeuGluMetAlaGluGluMetAsnLeuAspL 180
    :||||| |||:::       ::::::|||::::::|||       ::
10815 CTCTGAAATGAAAGCACTAGTGCCAATTGTACTCAA.......... 10779

180 euGlyLeuAspLysArgThrIleGluHisTyrValAspLeuGlnTyrVal 196
    :::::::       |||:::|||::::::||:::      :::
10778 .........ATTGAAGAATTCCCAGCAGTCATTACCTG 10749

197 ProGluProAspThrLeuHisAlaGlnIleSerArgLeuGluSerGlyCy 213
    ||:|||       :::|||:::
10748 TGGAGTAAGGAT.................................... 10737

213 sThrAlaThrValArgProGlyGlyLysLeuGluGlnLysArgTyrPheL 230
    :::::::|||       :::       ||||||
10736 ....GGTGTAGTCACGCCT...................TATTATC 10715

230 ysProGlnPheProValGlnLysValValLysGluLysLysGlnAspLeu 246
    :::::::
10714 AA.....................CGTGATTGGAAAGAT 10698

247 PheAspArgIleAla........................GlnVa 253
    ||||||||:||
10697 TTTGATAACGTTGCACAAAATGGTGGTGATAAAAGCGTTGTTAAGCAAGG 10648

253 lLeuGluAspSerValGluLysHisMetArgAlaAspValThrValGlyLys 270
    :::|||::::|||||| ||:||||||    :::::||||||
10647 TTTGGAAGATGCAGTTAAACGTCAATTAATGTGTGATGTACCTTACGGCG 10598

270 erPheLeuSerGlyGlyIleAspSerThrAlaIleAlaProLeuAlaLys 286
    :::|||:::::: ::::::|||       :::|||
10597 TATTATTATCAGGCGGGTTAGATTCATCTGTATTTCAGCAATCACCCAA 10548

287 ArgHisAsn. 289
    :::
10547 CAGTATGCAAAACGTCGTATTGAAGACGGTGGTAAAACCGAAGCATGGTG 10498

290 .ProAspLeuThrPheThrThrGlyPheGluArgGluArgGlyTyrSerG 306
    |||:::|||    ::::::|||:::::|||      :::|||
10497 GCCACAACTACACTCTTTCTCTGTAGGCCTA......AATGGTTCTCCAG 10454
```

```
DE   V. marinus PKS-like gene cluster encoding ORF6 to ORF9.
XX
KW   PKS pathway; polyunsaturated long chain fatty acid; plant; transgenic;
KW   polyketide-like synthesis; PUFA; dietary supplement; intravenous feeding;
KW   malnutrition; cooking oil; cooking fat; margarine;
KW   docosahexenoic acid production; eicosapentenoic acid production; ds.
XX
OS   Vibrio marinus.
XX
PN   WO200042195-A2.
XX
PD   20-JUL-2000.
XX
PF   14-JAN-2000; 2000WO-US00956.
XX
PR   14-JAN-1999;  99US-0231899.
XX
PA   (CALJ ) CALGENE LLC.
XX
PI   Facciotti D, Metz JG, Lassner M;
XX
DR   WPI; 2000-476063/41.
XX
PT   New DNA sequences encoding for polyketide (PK)-like synthesis pathway
PT   genes from Shewanella, Vibrio and Schizochtrium, useful for creating
PT   transgenic plants that express poly-unsaturated long chain fatty acids
PT   -
XX
PS   Example 1; Fig 5; 302pp; English.
XX
CC   This invention describes novel DNA sequences encoding for polyketide
CC   (PK)-like synthesis (PKS-like) pathway genes from Shewanella, Vibrio
CC   and schizochtrium. The nucleic acids are useful for isolating related
CC   molecules or in methods to detect organisms expressing the PKS-like
CC   genes. They are also useful for creating transgenic plants that express
CC   poly-unsaturated long chain fatty acids. The poly-unsaturated long chain
CC   fatty acids produced recombinantly are useful as dietary supplements for
CC   patients undergoing intravenous feeding or for preventing or treating
CC   malnutrition. The poly-unsaturated long chain fatty acids can also be
CC   incorporated into cooking oils, fats or margarine formulated so that in
CC   normal use the recipient receives a desired amount of poly-unsaturated
CC   long chain fatty acids. The nucleic acids are also useful in large scale
CC   production of docosahexenoic acid and eicosapentenoic acid, and for the
CC   modification of the fatty acid profile of host cells and edible plant
CC   tissues and/or plant parts. Transgenic production of polyunsaturated
CC   fatty acids in particular host cells allows quicker purification from
CC   natural sources such as fish or plants. This sequence represents a
CC   Vibrio marinus DNA fragment encoding the PKS gene cluster incorporating
CC   ORF6 to ORF9.
XX
SQ   Sequence 40138 BP; 11845 A; 7634 C; 8759 G; 11900 T; 0 other;
```

```
alignment_scores:
   Quality:    463.50              Length:    589
   Ratio:      1.525               Gaps:       20
   Percent Similarity: 51.613      Percent Identity: 25.127

alignment_block:
US-09-786-474-2 x AAA71520/rev
Align seg 1/1  to reverse of: AAA71520  from: 1 to: 40138

    1 MetCysGlyLeuLeuGlyIleLeuThrLeuLeuPheLeuLysGlyAsnAlaGluAlaPh 17
      |||||||:|:::::||||||||:||||||||
11291 ATGTGTCCAATATTCGGAATTCTAGACATCAAATCAGACATTAAA......    11247

   17 eValProAlaLeuGluArgAlaLeu.......ProCysMetArgHisA 31
      ||| :: :::|:||           :::|||||
11246 ....CCTCTTCGGACAACAGCACTAGAACTATCAAAGTTACTACGTCACC    11201

   31 rgGlyProAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGly 47
      ||||||||| ::: ||  :::|||
```

```
  306 luValAspValAlaAlaGluSerAlaAlaAlaIleGlyAlaIleHisIle 322
      ::::: ::|||  ::  |||  |||
10453 ATTTAGCCCTGCACAAAAGTGCAGATCACTCACTTAGGTACTATCCATCAT 10404

  323 ValLysIleValSerProGluGluTyrAlaAsnAlaIleProLysIleMe 339
      :::  :::||::  ||| | ||:::| ||
10403 TCAATTGAATTACTGTGCAAGAAGGTATCGACGCTTTACGCGACGTGAT    10354

  339 tTrpTyrLeuAsp.....AspProValAlaAspProSerLeuValProL 354
      ||||  ::||  |||    :: :::|||
10353 TTACCACATTGAAACCTACGATGTAACAACAATTCGTCGATCACCCCTA    10304

  354 euTyrPheValAlaAlaGluAlaArg...LysHisValLysValValLeu 369
      ::|||::||||::  :::  :::||:::|||||:|||||
10303 TGTACCTAATGGCACGTAAAATTAAAGCCATGGGGATTAAGATGGTGCTT 10254

  370 SerGlyGluGlyAlaAspProLeuPheGlyGlyTyrThrIleTyrLysGl 386
      |||||||||:||||||| |||    :: |||
10253 TCTGGTGAAGGTGCTGATGAACTCTTCGGTGGTTAC...TTGTAC.....  10212

  386 uProLeuSerLeuAlaProPheGluLysIleProSerProLeuArgLysG 403
          ||||:::  |||
10211 ..............TTCCACAAAGCGCCA................        10197

  403 lyLeuGlyLysSerLysValLeuProAspGlyMetLysGlyLysLysSer 419
      ::|||:|||||||||||||||||
10197 .................................                    10197

  420 LeuLeuGluArgGlySerMetThrMetGluGluArgTyrTyrTyrGlyAsnAl 436
      :::|||    :||
10196 ..........................AATGC                      10192

  436 aArgSerPheAsnPheGluGlnMetGlnArgValIleProTrpAlaLysA 453
      |::::||||| ||  ||||:::
10191 GAAAGAATTCCATGAAGAGACGGTACGTAAAGTA...........        10158

  453 rgGluTrpAspHisArgGluValThrArgAlaProIleTyrAlaGlnSerArg 469
      :::|| ::  ::  |||:::|||
10158 ..................                                   10158

  470 AsnPheAspProValAlaArgMetGlnHisLeuAspLeuPheThrTrpMe 486
      :: ::|| ::  |||
10157 ..........AATAAATTACATATGTTTGATTGCTT                 10132

  486 tArgGlyAspIleLeuValLysAlaAspLysIleAsnMetAlaAsnSerL 503
      ::||| |||  |||
10131 ACGT................GCAAACAAATCTATGGCAGCGTTGGGGCA    10100

  503 euGluLeuArgValProPheLeuAspLysGluValPheLysValAlaGlu 519
      ::||  ||||||||||||||
10099 TCGAAGCACGTGTTCCTTTCTTAGATAAAGAATTGTTGATGCGTCAATG   10050

  520 ThrIleProTyrAspLeuLysIleAlaAsnGly.....ThrThrLysTy 534
      :::||||||    |||
10049 CGTCTAAACCCTGAACTAAAAATGATCACGGGCACCCGCCATTGAGAAAAA 10000

  534 rAlaLeuArgArgAlaLeuGluGlnIleValProProHisValLeuHisA 551
      :::||||::||  :: |||  |
 9999 CATCATTGCGAGAGCATTGAAGATTTACTACCTGAAGAAATTGTATGGC   9950

  551 rgLysLysLeuGlyPhe 556
      ||::::|||  |||
 9949 GTCAAAAAGAACATTC                                     9933
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA71520

seq_documentation_block:
ID   AAA71520 standard; DNA; 40138 BP.
XX
AC   AAA71520;
XX
DT   11-DEC-2000  (first entry)
XX
```

```
11200 GCGGCCCAGATTGGTCAGTATCTATACAAGTGACAATGCTATTTTAGTT 11151

  48 PheAsnArgLeuSerIleIleAspIleAlaHisSerHisGlnProLeuAr  64
      :::||||||:::||:::||  ||          |||||||
11150 CATGAACGTCTTGCTATTGTCGATGTAAATAATGGTGCACAGCCCGCTA..  11103

  64 gTrpGlyProAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyG  81
      ::||||    ||||||             |||||||
11102 ..TATAATGAAGAAAAAGACACACGTACTTGCTGTTAATGGTG  11063

  81 luIleTyrAsnTyrValGlyLeuArgLysGluSerLeuGlyTyr  97
      |||||||:::||:::||||||:::||||  ::  ::
11062 AAATTTATAACCACAAAGATTTAAAGAAAAACGTG...AACGTAGACTTT  11016

  98 ThrPheAsnThrSerGlyAspGlyProIleValValGlyProheHisHi  114
      ||||||:::::|||||||||||::::::|||::: :::
11015 GAATTCCAAACAGATACAGATTGCGAAGTGATACTTGCGCTATATAAAGA  10966

 114 sTrpGlyGluSerValValGluHisLeuArgGlyMetPheGlyIleAlaI  131
      | :||  |||:::|||||||||::::|||||||||:::|
10965 AAAAGGCACACAATTTTAGATGACTAAATGGTATTTTTGCTTTCGCTT  10916

 131 leTrpAspThrLysGlyLysSerLeuPheLeuAlaArgAspGlnPheGly  147
      :::|||  |||  :::|||||||:::::||:::||
10915 TATATGACGAAACAGAGGACGCTTACCTGATGACCGTGACCATATCGGT  10866

 148 IleLysProLeuPheTyrAlaThrThrGluHisGlyThrValPhe...Se  163
      |||:::|| :::|:::|||||||||||:::|
10865 ATCATCCCTCTTTATACGGGCTATGATGAACACGTAACTTTATGTTGC  10816

 163 rSerGluLysLysThrIleLeuGluMetAlaGluGlnMetLeuAsnLeuAspL  180
      :||||||  ||| :::|:::||::||||||
10815 CTCTGAAATGAAAGCACTAGTGCCAATTGTACTTCAA...............  10779

 180 euGlyLeuAspLysArgThrIleGluHisTyrValAspLeuGlnTyrVal  196
      :::|||||||||
10778 ...............ATTGAAGAATTCCCAGCAGTCATTACCTG  10749

 197 ProGluProAspThrLeuHisAlaGlnIleSerArgLeuGluSerGlyCy  213
      :::|||
10748 TGGAGTAAGGAT...............  10737

 213 sThrAlaThrValArgProGlyGlyLysLeuGlnLysLysArgTyrPheL  230
      :::||||||  |||
10736 ...GGTGTAGTCACGCCT...............TATTATC  10715

 230 ysProGlnPheProValGlnLysValValLysGlyLysGluGlnAspLeu  246
      :||||||  |||||||
10714 AA...............CGTGATTGGAAAGAT  10698

 247 PheAspArgIleAla...............GlnVa  253
      ||||||:::|||     |||
10697 TTTGATAACGGTGCACAAAATGGTGGTAAACGTTGTGATAAAACGTCAAGG  10648

 253 lLeuGluAspSerValGlyHisMetArgAlaAspValThrValGlys  270
      :::|||||||:::|:::||  :::||||||||
10647 TTTGGAAGATAGCAGTGAACGTTAAACGTAATGTGTGATGTACCTTACGGCG  10598

 270 erPheLeuSerGlyGlyTyrIleAspSerThrAlaIleAlaProLeuAlaLys  286
      ||||||||||||:::||||||||::::::|:::|||
10597 TATTATTATCAGGCGGGTTAGATTCCATCTGTTATTTCAGCAATCACCCAA  10548

 287 ArgHisAsn...............  289
      :::::|
10547 CAGTATGCAAAACGTCTGTATTGAAGACGGTGGTAAAACCGAAGCATGGTG  10498

 290 .ProAspLeuThrPheThrThrGlyPheGluArgGluGlyTyrSerG  306
      ||||||||||  :::|:::||:::|
10497 GCCACAACTACACTCTTTCTCCGTAGGCCTA......AATGGTTCTCCAG  10454

 306 luValAspValAlaAlaValGluSerArgAlaAlaAlaIleGlyAlaGluHisIle  322
      :::|||  ||  :::|||:::|||  |||::: |||
10453 ATTTAGCCGCTGCACAAAAAGTGGCAGATCCACTTAGGTACTATCCATCAT  10404
```

```
 323 ValLysIleValSerProGluGluTyrAlaAsnAlaIleProLysIleMe  339
      :::||||:::     ::||||:::::|:::|
10403 TCAATTGAATTTACTGTCGAAGAGGTATCGACGCTTTACGCGACGTGAT  10354

 339 tTrpTyrLeuAsp...............AspProValAlaAspProSerLeuValProL  354
      :::||||||     |||     ::||||||
10353 TTACCACATTGAAACCTACGATGTAACAACAATTCGTCATCACCACCCCTA  10304

 354 euTyrPheAlaAlaAlaGluAlaarg...LysHisValLysValValLeu  369
      :||||:::|||||    ::    ::  ||||:|||||||||
10303 TGTACCTAATGGCACGTAAAATAAAGCCATGGGATTAAGATGGTGCTT  10254

 370 SerGlyGluGlyValAlaAspGluProheGlyGlyTyrThrIleTyrLysGl  386
      |||||||||||||||||||||||||:::||||||||
10253 TCTGGTGAAGGTGCTGATGAACTGTTCCGTGGTTAC...TTGTAC......  10212

 386 uProLeuSerLeuValAlaProheGluLysValProSerProLeuProArgLysG  403
      |||
10211 ...............TTCCACAAAGCGCCA...............  10197

 403 lyLeuGlyLysLeuSerLysValLeuProAspGlyMetLysGlyLysSer  419
      |||
10197 ...............  10197

 420 LeuLeuGluArgGlySerMetThrMetGluArgTyrTyrGlyAsnAl  436
      :::||||||
10196 ...............AATGC  10192

 436 aArgSerPheAsnPheGluGlnMetGlnArgValIleProAlaLysA  453
      :::|||:::||   :::||||||||||
10191 GAAAGAATTCCATGAAGAGACGGTACGTAAAGTA...............  10158

 453 rgGluTrpAspHisArgGluValThrAlaProIleTyrAlaGlnSerArg  469
      :::|||
10158 ...............  10158

 470 AsnPheAspProValAlaArgMetGlnHisLeuAspLeuPheThrTrpMe  486
      :::::|:::|
10157 ...............AATAAATTACATATGTTTGATTGCTT  10132

 486 tArgGlyAspIleLeuValLysAlaAspLysIleAsnMetAlaAsnSerL  503
      :||   |||||||:::||
10131 ACGT...............GCAAACAAATCTATGGCAGCTTGGGGCA  10100

 503 euGluLeuArgValProPheLeuAspLysGluValPheLysValAlaGlu  519
      :::|  |||||||||||||||||||||
10099 TCGAAGCACGTGTTCCTTTCTTAGATAAAGAAATTGTTGATGCGTCAATG  10050

 520 ThrIleProTyrAspLeuLysIleAlaAsnGly...............ThrThrLysTy  534
      |||     |||
10049 CGTCTAAACCCTGACTAAAATGATCACGGGCCACCGCATTGAGAAAAA  10000

 534 rAlaLeuArgArgAlaLeuGluGlnIleValProProHisValLeuHisA  551
      :::|||||:::|||||||||::::||
 9999 CATCATCCGAGAGGATCATTGAAGATTTACCTGAAGAAATTGTATGGC  9950

 551 rgLysLysLeuGlyPhe  556
      ||:::||
 9949 GTCAAAAAGAACAATTC  9933
```

```
KW   DHA industrial production; ds.
XX
OS   Moritella marina.
XX
PN   JP2001169780-A.
XX
PD   26-JUN-2001.
XX
PF   15-DEC-1999;   99JP-0356614.
XX
PR   15-DEC-1999;   99JP-0356614.
XX
PA   (KEIZ-) KEIZAI SANGYOSHO SANGYO GIJUTSU SOGO KEN.
XX
DR   WPI; 2001-592536/67.
XX
PT   Gene of a docosahexaenoic acid-producing microbe, used for industrial
PT   production of docosahexaenoic acid -
XX
PS   Claim 4; Page 10-24; 68pp; Japanese.
XX

CC   The present invention provides a DNA sequence encoding an icosapentaenoic
CC   acid-biosynthesising enzyme group-like protein group derived from a
CC   microbe having docosahexaenoic acid (DHA)productivity. The gene can be
CC   used for the industrial production of DHA. The present sequence is a
CC   coding sequence described in the exemplification of the invention.
XX
SQ   Sequence 41587 BP; 12323 A; 7885 C; 9126 G; 12253 T; 0 other;

alignment_scores:
        Quality:   463.50         Length:       589
        Ratio:     1.525          Gaps:          20
Percent Similarity:  51.613     Percent Identity:  25.127

alignment_block:
US-09-786-474-2 x AAI64984/rev

Align seg 1/1  to reverse of: AAI64984  from: 1  to: 41587

   1  MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPh  17
      |||||||||||||:||:||||||||   :: :::
3907  ATGTGCTCAAATTCGGAATTCTAGACATCAAATACAGACATTAAA.....  3863

  17  eValProAlaLeuGluArgAlaLeu.......ProCysMetArgHisA    31
      |||:|| |||||||||
3862  ....CCTCTTCCGCAACAAGCACTACGAACTATCAAAGTTACTACGTCACC  3817

  31  rgGlyProAspAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGly  47
      ||||||||  :::  :|||| ::|||
3816  GCGGCCCAGATTGGTCAGGTATCTATACAAGTGACAATGCTATTTAGTT   3767

  48  PheAsnArgLeuSerIleIleAspIleAlaHisSerHisGlnProLeuAr  64
      ::::::|||||:|:::|||| :: :::|||| |||||||
3766  CATGAACGTCTTGCTATTGTCGATGTAAATAATGGTGCACAGCCGCCTA.. 3719

  64  gTrpGlyProAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyG  81
      :||| |||||||| ||||||
3718  ........TATAATGAAGAAAAGACACACGTACTTGCTGTGTTAATGGTG  3679

  81  luIleTyrAsnTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyr  97
      |||||||||||||:::|||  :::::
3678  AAATTTATAACCACACAAGATTTAAAGAAAACGCTG...AACGTAGACTTT 3632

  98  ThrPheAsnThrSerGlyAspGluProIleValGlyPheHisHi     114
      ||||||||||:::::||||||||||||||
3631  GAATTCCAAACAGAATCAGATTGCGAAGTGATACTTGCGCTATATAAAGA  3582

 114  sTrpGlyGluSerValValGluHisLeuArgGlyMetPheGlyIleAlaI 131
      ||| ::: :::::|||:::|||:::||::::|||
3581  AAAAGGCACACAATTTTAGATGATCTAAATGGTATTTTGCTTTCCGCTT   3532

 131  leTrpAspThrLysGluLysSerLeuPheLeuAlaArgAspGlnPheGly 147


      :::|||
3531  TATATGACGAAACAGAGGACGTTACCTGATTGGCCGTGACCATATCGGT   3482

 148  IleLysProLeuPheTyrAlaThrThrGlnHisGlyThrValPhe...Se 163
      |||  ||||||| ::: |||||||::::
3481  ATCATCCCTCTTTATACGGCTATGATGACACGGTAACTTTTATGTTGC    3432

 163  rSerGluLysThrIleLeuGluMetAlaGluHisMetAlaGluMetAsnLeuAspL 180
      :: :::
3431  CTCTGAAATGAAAGCACTAGTGCCAATTGTACTCAA.........       3395

 180  euGlyLeuAspLysArgThrIleGluHisTyrValAspLeuGlnTyrVal 196
      ||||||||:::::|||
3394  ........ATTGAAGAATTCCCAGCAGTCATTACCTG              3365

 197  ProGluProAspThrLeuHisAlaGlnIleSerArgLeuGluSerGlyCy 213
      ::: |||
3364  TGGAGTAAGGAT...........                            3353

 213  sThrAlaThrValArgProGlyGlyLeuGluGlnLysArgTyrPheL    230
      :::::||| |||
3352  ...GGTGTAGTCACGCCT............         TATTATC     3331

 230  ysProGlnPheProValGlnLysValValLysGlyValLysGluGlnAspLeu 246
      :: :
3330  AA...........CGTGATTGGAAAGAT                       3314

 247  PheAspArgIleAla...........GlnVa                    253
      |||||||::::|||
3313  TTTGATAACGTTGCACAAAATGGTGGTGATAAAAGCGTTGTTAAGCAAGG 3264

 253  lLeuGluAspSerValGlyLysHisMetArgAlaAspValThrValGlyS 270
      :::::|||::::||||| |||
3263  TTTGGAAGATGCAGTTAAACGTCAATTAATGTGTGATGTACCTTACGGCG 3214

 270  erPheLeuSerGlyIleTyrIleAspSerThrAlaIleAlaProLeuAlaLys 286
      ::::||||||||:::::||::|||:::::||:::::
3213  TATTATTATTCAGGGATTTAGATTCATCTGTTATTTCAGCAATCACCCAA 3164

 287  ArgHisAsn.                                         289
      ::::::
3163  CAGTATGCAAAACGTCGTATTGAAGACGGTGGTAAAACCGAAGCAGTGTG 3114

 290  .ProAspLeuThrPheThrPheThrGlyPheGlyArgGlyGlyTyrSerG 306
      :::||||| ::||||| :::
3113  GCCCAACTACACTCTTTCTGTGTAGGCCTA......AATGGTTCTCCAG   3070

 306  luValAspValAlaAlaGluSerAlaAlaIleAlaGlyHisGluHisIle 322
      ::::|||||| |||
3069  ATTTAGCCGCTGCACAAAAGTGGCAGATCACTTAGGTACTATCCATCAT  3020

 323  ValLysIleValSerProGluGlyTyrAlaAlaAsnAlaIleProLysIleMe 339
      ::|||| |||| :::::
3019  TCAATTGAATTACTGTGCCAAGAGGTATCGACGCTTTACGCGACGTGAT   2970

 339  tTrpThrTyrLeuAsp......AspProValAlaAspProSerLeuValProL 354
      :::||| :::::
2969  TTACCACATTGAAACCTACGATGTAACACAATTCGTGCATCAACCCCTA   2920

 354  euTyrPheValAlaAlaGluAlaArg...LysHisValLysValValLeu 369
      :::|||||||||| ::::||:||||||
2919  TGTACCTATTGGACGTAAAATTAAAGCCATGGGGATTAAGATGGTGCTT   2870

 370  SerGlyGluGlyAlaAspAlaAspLeuPheGlyGlyTyrThrIleTyrLysGl 386
      |||||||||||||||||::::||
2869  TCTGGTGAAGGTGCTGATGAACCTTCGGTGGTTAC...TTGTAC......  2828

 386  uProLeuSerLeuAlaProPheGluLysIleProSerProLeuArgLysG 403
      :::::|||::::
2827  ...........TTCCAAAGCGCCA.                          2813

 403  lyLeuGlyLysLeuSerLysValLeuProAspGlyMetLysGlyLysSer 419
```

us-09-786-474-2.p2n.rng

```
2813 ............................................................ 2813

 420 LeuLeuGluArgGlySerMetThrMetGluGluArgTyrTyrGlyAsnAl          436
                                                 |||||
2812 .................................................AATGC      2808

 436 aArgSerPheAsnPheGluGlnMetGlnArgValIleProTrpAlaLysA          453
     |::::::||::: ||| :::::::|||
2807 GAAAGAATTCCATGAAGAGACGGTACGTAAAGTA.                         2774

 453 rgGluTrpAspHisArgGluValThrAlaProIleTyrAlaGlnSerArg          469

2774 ............................................................ 2774

 470 AsnPheAspProValAlaArgMetGlnHisLeuAspPheThrTrpMe             486
                          ::: |||  :::|||
2773 ...................AATAAATTACATATGTTTGATTGCTT               2748

 486 tArgGlyAspIleLeuValLysAlaAspLysIleAsnMetAlaAsnSerL          503
     :|||           |||::::|||    ||| :::::
2747 ACGT.................GCAAACAAATCTATGGCAGCTTGGGGCA           2716

 503 euGluLeuArgValProPheLeuAspLysLysGluValPheLysValAlaGlu       519
     ::|||       |||:||||||||||||||||        :::::
2715 TCGAAGCACGTGTTCCTTTCTTAGATAAAGAATTGTTGATGCGTCAATG          2666

 520 ThrIleProTyrAspLeuLysLysIleAlaAsnGly.......ThrThrLysTy      534
     :::  :|||||| :::||||            ::|||
2665 CGTCTAAACCCTGAACTAAAAATGATCACGGGCGACCGCATTGAGAAAAA         2616

 534 rAlaLeuArgArgAlaLeuGluGluIleValProProHisValLeuHisA          551
     :::||||:||::||||  ::::||||    :::::::::|
2615 CATCATTCGAGAAGCATTTGAAGATTACTACCTGAAGAAATTGTATGGC          2566

 551 rgGlyLysLeuGlyPhe  556
     ||::::|||       |||
2565 GTCAAAAGAACAATTC  2549
```

```
OM of: US-09-786-474-2 to: EST:*    out_format : pfs

Date: Sep 20, 2002  7:47 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO_spool/US09786474/runat_18092002_132910_5427/app_query.fasta_1.706
-DB-EST -QFMT=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-LIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
-INLEN=0 -MAXLEN=2000000000 -USER=US09786474_@CGN1_1_3007
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-786-474-2
Query length: 640
Database: EST:*
Database sequences: 13736207
Database length: -1841457050
Search time (sec): 1891.720000
```

```
gb_est1:AL559639        +    234.00   423.44   2.4e-14   | AL559639 AL559639 LTI_FL011_
gb_est1:AW773806        +    233.50   425.86   1.8e-14   | AW773806 EST332792 KV3 Medic
gb_est2:BI417222        +    232.50   426.04   1.7e-14   | BI417222 LjNEST16d5r Lotus j
gb_est2:BI418097        +    232.50   424.54   2.1e-14   | BI418097 LjNEST24e8r Lotus j
gb_est1:AU083977        +    232.00   426.48   1.7e-14   | AU083977 AU083977 Cryptomer1
```

seq_name: gb_est2:BI308442

```
seq_documentation_block:
LOCUS       BI308442                802 bp    mRNA    linear   EST 20-JUL-2001
DEFINITION  EST529852 GPOD Medicago truncatula cDNA clone pGPOD-5L8 5' end,
            mRNA sequence.
ACCESSION   BI308442
VERSION     BI308442.1  GI:14982769
KEYWORDS    EST.
SOURCE      barrel medic.
ORGANISM    Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 802)
AUTHORS     Grusak,M.A., Samac,D.A., Town,C.D., Van Aken,S., Utterback,T., Cho
            ,J. and Fraser,C.M.
TITLE       ESTs from developing reproductive tissues of Medicago truncatula
JOURNAL     Unpublished (2001)
COMMENT     Contact: Michael A. Grusak
            USDA/ARS Children's Nutrition Research Center
            Baylor College of Medicine
            1100 Bates Street, Houston, TX 77030-2600, USA
            Tel: 713-798-7044
            Fax: 713-798-7078
            Email: mgrusak@bcm.tmc.edu
            B395297e
            TIGR sequence name: MTOAP64TK
            More information is available at: www.medicago.org
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1..802
                     /organism="Medicago truncatula"
                     /cultivar="A17"
                     /db_xref="taxon:3880"
                     /clone="pGPOD-5L8"
                     /clone_lib="GPOD"
                     /tissue_type="immature pod walls"
                     /dev_stage="Immature pods, ranging in age from 15 to 30
                     days after pollination"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Immature pods, ranging in age from 15 to 30 days
                     after pollination, were collected from greenhouse-grown
                     plants. At harvest, seeds were removed from pods and
                     isolated pod walls were collected and immediately frozen
                     in liquid nitrogen. Pod walls were pooled for mRNA
                     extraction. cDNA was prepared from polyA+ enriched RNA.
                     The cDNA was directionally ligated into the Unizap XR
                     vector from Stratagene and packaged using Gigapack III
                     Gold packaging extracts. Plasmids containing cDNA inserts
                     were excised from the recombinant lambda-Zap phage using
                     Ex-assist helper phage and propagated in XLOLR cells."
BASE COUNT      200 a    157 c    196 g    249 t
ORIGIN
```

```
alignment_scores:
            Quality:       310.00          Length:       292
            Ratio:           1.813           Gaps:          7
Percent Similarity:  58.562    Percent Identity:  29.110

alignment_block:
US-09-786-474-2 x BI308442

Align seq 1/1 to: BI308442     from: 1   to: 802
```

```
score_list:
Sequence                  Strd  Orig     ZScore   EScore   Len |  Documentation
gb_est2:BI308442           +   310.00   570.07   1.7e-22  802  | BI308442 EST529852 GPOD Medicag
gb_est2:BG366239           +   284.00   520.18   1.0e-19  809  | BG366239 HVSME10006F21f Hordeum
gb_est1:AW720554           +   275.00   502.12   1.0e-18  869  | AW720554 LjNEST10p9rc Lotus jap
gb_est1:AL538252           +   264.00   480.01   1.7e-17  952  | AL538252 AL538252 LTI_FL013_FB1
gb_est2:BF622268           +   263.00   479.35   1.9e-17  854  | BF622268 HVSMEa0002G16f Hordeum
gb_est1:AL555468           +   259.00   470.74   5.7e-17  927  | AL555468 LTI_NFL006_PI
gb_est1:BI179230           +   255.50   467.16   9.0e-17  707  | BI179230 EST520175 cSTE Solanum
gb_est2:BI269282           -   254.00   464.42   1.3e-16  699  | BI269282 NF006E11R1F1086 Irrad
gb_est1:AL515757           +   253.50   458.32   2.8e-16 1092  | AL515757 AL515757 LTI_NFL011_N
gb_est2:BM457834           +   253.00   459.98   2.2e-16  870  | BM457834 AGENCOURT_6412078 NIH
gb_est2:BE325921           +   249.00   455.25   4.1e-16  675  | BE325921 NF084F12S1F1096 Devel
gb_est2:BF618750           +   249.00   453.46   5.1e-16  788  | BF618750 HVSMEc0007K11f Hordeum
gb_est1:AV925031           +   243.50   446.65   1.2e-15  571  | AV925031 K.  Sato unpub
gb_est2:BE405536           +   243.50   446.00   1.4e-15  604  | BE405536 WHE1212_E01_J02ZS Whea
gb_est1:AV916129           +   243.50   445.90   1.4e-15  609  | AV916129 AV916129 K.  Sato unput
gb_est1:AV924234           +   243.50   445.66   1.4e-15  622  | AV924234 AV924234 K.  Sato unput
gb_est1:AV923285           +   243.50   445.62   1.4e-15  624  | AV923285 AV923285 K.  Sato unput
gb_est1:AV924053           +   243.50   445.06   1.5e-15  655  | AV924053 AV924053 K.  Sato unput
gb_est2:BG127495           +   243.50   443.42   1.9e-15  755  | BG127495 EST473141 tomato shoot
gb_est2:BG310222           .   243.50   442.17   2.2e-15  842  | BG310222 HVSMEc0016I16f Hordeum
gb_est2:BE470565           +   242.50   445.16   1.5e-15  550  | BE470565 WHE0261_D02_D02ZS Whea
gb_est2:BE490756           +   242.50   444.87   1.6e-15  564  | BE490756 WHE0368_G11_M22ZS Whea
gb_est2:BE630423           +   241.50   440.19   2.8e-15  717  | BE630423 HVSMEb0009F05f Hordeum
gb_est2:BF258379           +   241.50   439.67   3.0e-15  750  | BF258379 HVSMEf0015J02f Hordeum
gb_est2:BF631030           +   241.50   439.40   3.2e-15  768  | BF631030 HVSMEb0014M07f Hordeum
gb_est2:BF626813           +   241.50   437.63   4.0e-15  895  | BF626813 HVSMEb0001G12f Hordeum
gb_est2:BG468478           +   241.00   437.14   4.2e-15  860  | BG468478 6025104S2F1 NIH_MGC_15
gb_est1:AV927055           +   240.50   441.23   2.5e-15  555  | AV927055 AV927055 K.  Sato unput
gb_est2:BE628160           +   238.50   438.44   3.6e-15  507  | BE628160 HVSMEb0003E18f Hordeum
gb_est2:BG882033           +   238.50   437.82   3.9e-15  535  | BG882033 sae95b10.y1 Gm-c1065 G
gb_est2:BG452857           +   238.50   435.44   5.2e-15  658  | BG452857 NF081E08L1F1056 Devel
gb_est2:BG299830           +   237.50   431.53   8.6e-15  782  | BG299830 HVSMEa0021P24f Hordeum
gb_est1:AL536791           +   237.00   429.09   1.2e-14  889  | AL536791 AL536791 LTI_FL013_FB1
gb_est2:BF474366           +   236.00   433.79   6.5e-15  501  | BF474366 WHE0843_B11_D21ZS Whea
gb_est2:BE404585           +   236.00   433.57   6.7e-15  511  | BE404585 WHE0445_A06_A11ZS Whea
gb_est2:BE591022           +   236.00   433.57   6.7e-15  511  | BE591022 WHE1655-1658_J04_J04ZS
gb_est2:BE340158           +   236.00   433.34   6.9e-15  521  | BE340158 EST344227 potato stolo
gb_est2:BG045081           +   236.00   430.65   9.7e-15  658  | BG045081 saa38c02.y1 Gm-c1059 G
gb_est2:BI719579           +   235.50   430.01   1.1e-14  640  | BI719579 1031044E03.y1 C. reinh
gb_est2:BI527119           +   234.50   428.82   1.2e-14  601  | BI527119 1024078D07.y1 C. reinh
```

```
seq_documentation_block:
LOCUS          BG366239                 809 bp    mRNA            linear   EST 22-OCT-2001
DEFINITION     HVSMEl0006F21f Hordeum vulgare 20 DAP spike EST library HVCDNA0010
               (20 DAP) Hordeum vulgare cDNA clone HVSMEl0006F21f, mRNA sequence.
ACCESSION      BG366239
VERSION        BG366239.2  GI:16324128
KEYWORDS       EST.
SOURCE         barley.
ORGANISM       Hordeum vulgare
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
               ; Triticeae; Hordeum.
REFERENCE      1  (bases 1 to 809)
AUTHORS        Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
               ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton
               ,R.D., Close,S.J., Oates,R. and Main,D.
TITLE          Development of a genetically and physically anchored EST resource
               for barley genomics: Morex 20 DAP spike cDNA library
JOURNAL        Unpublished (2001)
COMMENT        On Mar 8, 2001 this sequence version replaced gi:13255338.
               Contact: Wing RA
               Clemson University Genomics Institute
               Clemson University
               100 Jordan Hall, Clemson, SC 29634, USA
               Tel: 864 656 7288
               Fax: 864 656 4293
               Email: rwing@clemson.edu
               Total hq bases = 585
               Seq primer: AATTAACCCTCACTAAAGGG
               High quality sequence start: 6
               High quality sequence stop: 730.
FEATURES             Location/Qualifiers
     source          1..809
                     /organism="Hordeum vulgare"
                     /cultivar="Morex"
                     /db_xref="taxon:4513"
                     /clone="HVSMEl0006F21f"
                     /clone_lib="Hordeum vulgare 20 DAP spike EST library
                     HVCDNA0010 (20 DAP)"
                     /tissue_type="20 DAP spike"
                     /lab_host="SOLR"
                     /note="Vector: lambdaZAP; Site_1: EcoRl; Site_2: XhoI;
                     Plants were grown in the greenhouse at the University of
                     California, Riverside (Fenton, SJ Close, TJ Close). Whole
                     spikes with awns trimmed were collected at 20 DAP (Fenton
                     ). Total RNA was prepared, poly(A) RNA was purified, one
                     primary unamplified cDNA library was made, and 1 million
                     pfu were in vivo excised to give pBluescript SK(-) cDNA
                     phagemids in the TJ Close lab at the University of
                     California, Riverside (Choi). Phagemids were plated and
                     picked at the Clemson University Genomics Institute (CUGI)
                     (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA
                     preparations, DNA sequencing and sequence analysis were
                     performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates
                     , Rambo, Main). The sequence has been trimmed to remove
                     vector sequence and contains a minimum of 100 bases of
                     phred value 20 or above. For more details on library
                     preparation and sequence analysis see
                     http://www.genome.clemson.edu/projects/barley. To order
                     this clone see http://www.genome.clemson.edu/orders Also
                     see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                     Genetically and physically anchored EST resources for
                     barley genomics. Barley Genetics Newsletter 31:29-30.
                     (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT       173 a     231 g     182 t     1 others
ORIGIN           222 c
```

```
alignment_scores:
         Quality:  284.00          Length:      288
         Ratio:     1.701          Gaps:         13
Percent Similarity: 57.986    Percent Identity: 31.597
```

```
  1 MetCysGlyLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPh     17
    |||||||||||||::::::|||          |||:|||||
 33 ATGTGCGGAATCTTAGCAGGTTGGGTGGTGGTCGACAATTCTCAGGCCAA  82

 17 eValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA  34
    :::::|||          :||  :||  |||||||||||||||||||
 83 ACGGCGTCGCATCATCGAATTGTCTCGTAGGTTGCCACATAGAGGTCCTG 132

 34 spAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGlyPheAsnArg  50
    ||  :::|||              ||||  :::::|  :::|||
133 ATTGGAGTGGTTTGCATTGTCATCAAGATTGTTACCTCGCTCATCAACGT 182

 51 LeuSerIleIleAspIleAlaHisSerHisGlnProLeuArgTrpGlyPr  67
    |||::::|||  ::  :||||          ||||||
183 CTTTGCTATGTGACCCTACTTCCGGAGATCAACCTCTT............ 221

 67 oAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyGluIleTyrA  84
       ::||||||  :||  |||  :||:  ||:||||||||||||:|||||
222 .TACAACTATGACAAGACCGTTATTGTCACCGTTAATGGGAGATATACA  270

 84 snTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyrThrPheAsn 100
    ||::          :::||||||||||::::||||          :::
271 ACCAACAAGAAATTGAGCAGCAGAAACTGAAA.....TCTCATCAATTCGA 314

101 ThrSerGlyAspGlyProIleValValGlyPheHisHisHisTrpGlyGl 117
    |||::::|||  |||  |||  :||  ||||||::::::::||||||
315 ACTGGTAGTGACTGTGAAGTCATTGCTCATCTTATGAAGAATATGGTGA  364

117 uSerValValGluHisLeuArgGlyMetPheGlyIleAlaIleTrpAspT 134
    ||:::  |||::  ||||||::  |||          ::|||||||||
365 AGAGTTGTTGAATGCTGGATGGGGATGTTCCATTTGTGCTTCTTGATA  414

134 hrLysGluLysSerLeuPheLeuLeuAlaArgAspGlnPheGlyIleLeu 150
    |||::::|||::||||||::||  ||||          ::|||||  |||
415 CTCGAGATAAAAGTTTCATTGCTGCTCGTGATGCTATTGGCATTACCCCT  464

151 LeuPheTyrAlaThrThrGluHisGlyThrVal...PheSerSerGluLy 166
       :::          ::  :|||||||          |||::|||||||
465 CTTTACTTGGGTTGGGCCATGATGGATCAATATGGTTGCGTCTGAAAT   514

166 sLysThrIleLeuGluMetAlaGluGluMetAsnLeuAsnLeuGlyLeuA 183
    ||::||||          :|||  :||||||          ::|||||
515 GAAAGCTCTGATTGACGATTGTGAG....................      539

183 spLysArgThrIleGluHisTyrValAspLeuGlnTyrValProGluPro 199
       |||::||          ::|||||  :|||||||||||||
540 ..........CAATTCATATCTTTTCCT                      557

200 AspThrLeuHisAlaGlnIleSerArgLeuGluSerGlyCysThrAlaTh 216
       :::          ::|||  :::|||||:  :::|||||  ::
558 ..............................CCAGGCGCATATATATTC   574

216 rValArgProGlyGlyLysGluGlnLysArgTyrPheLysProGlnProP 233
       :|||||  :|||  ||||||||::|||          :||||
575 CAGCAAACAGGGAGGA......TTAAGAAGGTGGTATAATCCACCAT  615

233 heProValGlnLysValValLysGlyLysGluGlnAspLeuPheAspArg 249
    |||::::::|||          ::||||||||||||||||||||
616 GGTTCAGAGAGAAAATTCCATCAACACGCTATGATCCAACGGTT.....  659

250 IleAlaGlnValLeuGluAspSerValGluLysHisMetArgAlaAspVa 266
       :  :||||          :::||||||||::||||::::|||:||||
660 TTGCGTGAGGGCCTTTGAGAGGCGTGTAGTTAAGAGAATGATCACTGATGT 709

266 lThrValGlySerPheLeuSerPheLeuGlyIleAspSerThrAlaIleAlaP 283
    |||||||||||||::::||||||||||::|||::|::|:|||||:|||
710 ACCCTTTGGAGTTCTTTTGTCCCGGAGACTGGACTTCACTTGTTGCTG 759

283 roLeuAlaLysArgHisAsnProAsp 291
       ::||||:|||:|||  |||:|||
760 CAGTGGCCAATCGATATTTGGCTGAT 785
```

```
seq_name: gb_est2:BG366239
```

```
alignment_block:
US-09-786-474-2 x BG366239

Align seg 1/1  to: BG366239    from: 1 to: 809

  1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGlu.....   15
    |||||||||:::||:::||||||
 80 ATGTGCGGCACTACTGGCGGTGCTGGGGTGCCGGGCAGAGTCCAGGGGAA  129

 16 ...AlaPheValProAlaLeuGluArgAlaLeuProCysMetArgHisA    31
    :::|||   ||| :|||||||
130 GAGGGTGCATCTGCTCGAGCTCCGCGCAGGCTC.......AAGCACC     170

 31 rgGlyProAspAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGly   47
    |||||||||      :::||||||
171 GCGGGCCCGGACTGGAGCGGCCTGCACCAGGTCGCCGACAACTACCTCTGC  220

 48 PheAsnArgLeuSerIleIleAspIleAlaHisSerHisGlnProLeuAr   64
    ||||||||||||||||| ||| |||||||
221 CACCAGCGCCTCGCCATCGACCCGGCCTCCGGCGACCAGCCGCTC..     268

 64 gTrpGlyProAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyG   81
    :||||
269 .........TACAACGAGGACAAGTCCGTTGCTGTTGCCGTCAACGGGG    308

 81 luIleTyrAsnTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyr   97
    ||||||||||:::||||||||:::::||||   |||
309 AGATCTACAACCATGAGAGCTCCGGAACGGCTCCCC.....GGCCAC      352

 98 ThrPheAsnThrSerGlyAspGlyGluProIleValGlyPheHisHi    114
    ||:::||||||:::||||     :::::|||
353 AGGTTCCGGACCGGAAGTGACTGCGAGGTCATCGCCCATCGTATGAGGA   402

114 sTrpGlyGluSerValValGluValGlyMetPheGlyIleAlaI      131
    :   ||||||||||||||:   |||
403 ACATGGAGAAAGCTTCATTGACATGTTGGATGGTGTTTTCGTTCGTGT    452

131 leTrpAspThrLysGluLysSerLeuPheLeuAlaArgAspGlnPheGly  147
    ::|||||||||  |||::::: ||||||||   :::|||
453 TGCTTGACGCAGAGATCACAGCTTCATTGCTGTCGTGATGCCATTGGT    502

148 IleLysProLeuPheTyrAlaThrThrGluHisGlyThrValPhe...Se  163
    :::|||||||  ||||||||||    ||
503 GTCACGCCTCTACATTGGCTGGGAATCAGTGGTTCAGTGTGGATATC     552

163 rSerGluLysLysThrThrLeuLeuMetAlaGluMetAsnLeuAspL    180
    |||||| |||
553 TTCGGAGATGAAA......................               565

180 euGlyLeuAspLysArgThrIleGluHisTyrValAspLeuGlnTyrVal  196
    :||||:::                   ||||||
566 ..GGACTAAACGAT..GATTGTGAGCACTTC.......GAGATCTTC     601

197 ProGluProAspThrLeuHisAlaGlnIleSerArgLeuGluSerGlyCy  213
    |||   ||||||| :::               ||
602 CCA...CCTGTAATCTTTACTCCAGC...........AAAGAANAGTG    636

213 sThrAlaThrValArgProGlyGlyLysGluGlnLysArgTyrPheL    230
    |
637 CTTC.....................AAGAGATGGTATA             653

230 ysProGlnPheProValGlnLysValValLysGlyValGlyLysAspLeu 246
    :::||                     ::|||
654 ACCCTTCCTTGG............TTCTCGAGGGGTCATCCCTCGGTTCCC 691

247 PheAsp......ArgIleAlaGlnValLeuGluAspSerValGluLysHi  261
    :::|||     |||   ||||  :::||
692 TATGACCCACTACGTCCAGATGCGCATTTGAAAAGGCTGTCATCAAGAA   741

261 sMetArgAlaAspValThrValIleGlySerPheLeuSerGlyIleAspS  278
    :::||||||||:::|||||||||
742 GCTCATGACAGAAGATGTTCCATTCGGTTCCTCTTCTGGTGGCCTGGACT  791
```

```
278 erThrAlaIleAla   282
    ||||||||||
792 CATCACTGGTGGCA   805
```

```
seq_name: gb_est1:AW720554

seq_documentation_block:
LOCUS       AW720554              869 bp    mRNA    linear   EST 19-APR-2000
DEFINITION  LjNEST10b9rc Lotus japonicus nodule library 5 and 7 week-old Lotus
            japonicus cDNA 5', mRNA sequence.
ACCESSION   AW720554
VERSION     AW720554.1  GI:7615100
KEYWORDS    EST.
SOURCE      Lotus japonicus
  ORGANISM  Lotus japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 869)
  AUTHORS   Colebatch,G., Freund,S., Trevaskis,B and Udvardi,M.
  TITLE     Lotus japonicus root nodule ESTs: tools for functional genomics
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Udvardi MK
            Molecular Plant Nutrition
            Max Planck Institute of Molecular Plant Physiology
            Am Muehlenberg 1, 14476 Golm, Germany
            Fax: 49 331 567 8250
            Email: udvardi@mpimp-golm.mpg.de
            Seq primer: T7
            High quality sequence stop: 869.
FEATURES             Location/Qualifiers
     source          1..869
                     /organism="Lotus japonicus"
                     /cultivar="Gifu (B-129)"
                     /db_xref="taxon:34305"
                     /clone_lib="Lotus japonicus nodule library 5 and 7
                     week-old"
                     /dev_stage="5 and 7 week-old plants"
                     /note="Organ: Nodule; Vector: pSPORT1; Site_1: SalI;
                     Site_2: NotI; The library was prepared using mRNA
                     extracted from nodules of 5 and 7 week-old Lotus plants.
                     Nodules were induced by, and contained Mesorhizobium
                     strain R7A."
BASE COUNT      203 a     182 c     191 g     293 t
ORIGIN
```

```
alignment_scores:
            Quality:     275.00          Length:     286
            Ratio:         1.637         Gaps:         8
Percent Similarity:       58.741    Percent Identity:   28.322

alignment_block:
US-09-786-474-2 x AW720554

Align seg 1/1  to: AW720554    from: 1 to: 869

  1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAla...  16
    |||||||||||||||:::||| |||:::||||||
140 ATGTGCGGCATACTTGCTGCCTGCTTGGTTGTTCTGATGACTCTCAAGCCAA 189

 17 ......PheValProAlaLeuGluArgAlaLeuProCysMetArgHisA   31
          |||||||||| ||| ||| :::|||
190 AAGGGTTGCGTCGTCCTTGAGCTTTCCGCAGATTG.........AAGCACC  230

 31 rgGlyProAspAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGly   47
    ||:::|||    |||||||
231 GTGGCCCTGACGGAGTGGGCTCCACCAACATGGTGATAACTTTTTGGCT   280

 48 PheAsnArgLeuSerIleIleAspIleAlaHisSerHisGlnProLeuAr   64
    ||||||||      ::: |||
742 PheAsnArgLeuSerIleIleAspIleAlaHisSerHisGlnProLeuAr
```

TITLE   Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   Contact: Genoscope
           Genoscope - Centre National de Sequencage
           BP 191 91006 EVRY cedex - France
           Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES   Location/Qualifiers
     source   1..952
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DF027Yi13"
                     /clone_lib="LTI_FL013_FBrn1"
                     /dev_stage="pooled tissue from post conception fetuses (20
                     week, 24 week and 26 week)"
                     /lab_host="DH10B"
                     /note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
                     cDNA was primed with a NotI-oligo(dT) primer. Five prime
                     end enriched, double-stranded cDNA was digested with Not I
                     and cloned into the Not I and Eco RV sites of the
                     pCMVSPORT 6 vector. Library was constructed by Life
                     Technologies. Contact : Feng Liang Life Technologies, a
                     division of Invitrogen 9800 Medical Center Drive Rockville
                     , Maryland 20850, USA Fax : (1) 301 610 8371 Email :
                     filang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT   260 a   199 c   228 g   262 t   3 others
ORIGIN

alignment_scores:
      Quality: 264.00   Length: 325
      Ratio: 1.509   Gaps: 15
Percent Similarity: 53.846   Percent Identity: 27.385

alignment_block:
US-09-786-474-2 x AL538252

Align seg 1/1   to: AL538252   from: 1   to: 952

 19 ProAlaLeuArgAlaLeuProCysMetArg.......HisArgGlyPr   33
     |||||| :::::||| :::||||:||||| ||||||||||
 84 CCTGCTTACGCCCAGTGTCTGAGTGCTATGAAGATTGCACCAGAGGTCC 133

 33 oAspAspAlaGlyThrTrpHisAsp...........AlaAspAlaAlaP   46
     : ||||| ::::::::|||||||| |||||||||||
134 A...GATGGCATTCCGTTTGAGATGTCAATGGATACACCAACTGCTGCT 180

 46 heGlyPheAsnArgLeuSerIleIleAspIleAlaHisSerGlnInPro   62
     ||||||||::: |||||||||||:|| ::: ||||| |||||||
181 TTGGATTTCACCGGTTGGCCGGTAGTTGACCCCGTGTTTGGAATGCAGCCA 230

 63 LeuArg.....................TrpGlyProAlaAspGluProAs   72
     ::|||| |||
231 ATTCGAGTGAAGAAATATCCGTATTTGTG.   260

 72 pArgTyrAlaMetThrProAsnGlyGlyLeuTyrAsnTyrValGluLeuA   89
     ||||||||||||||||||||||||||||| :::::|||||||||||
261 .........CCTCGTTACAATGTGAAATCTACAACCATAAGAAGAATGC 300

 89 rgLysGluLeuSerAspLeuGlyTyrThrPheAsnThrSerGlyAspGly 105
     ::: :|||||||||| ::: :::||||||| ||||||||
301 AACAGCATTTTGAA.........TTTGAATACCAGACCAAAGTGGATGGT 341

106 GluProIleValValGlyPheHisHisTrpGly...GluSerValValGl 121
     || ||||||||||||||| |||||||||| |||||||||||||
342 GAGATAATCCTTCATCTTTATGACAAGGAGGAATTGAGCAAACAATTTG 391

121 uHisLeuArgGlyMetPheGlyIleAlaIleTrpAspThrLysGluLysS 138
     |:::|:::: :||:||||:: ::: |||||||:||| ::: :::|||
392 TATGTTGGATGGGTGTGTTTGCATTTGTTTTACTGGATACTGCCAATAAGA 441

138 erLeuPheLeuAlaArgAspArgGlnPheGlyIleLysProLeuPheTyrAla 154
     :::|||||||| ::||||||| |||||||||||||||||||

```
442 AAGTGTTCCTGGGTAGAGATACATATGGAGTCAGACCTTTGTTTAAAAGCA 491
        ||||| |||
155 ThrThrGluHisGly...ThrValPheSerSerGluLysLysThrIleLe 170

492 ATGACAGAAGATGATTTTTGGCTGTATGTTCAGAAGCTAAAGGTCTTGT 541
        ||| :::::: || |     ||| :::::
170 uGluMetAlaGluGluMetAsnLeuAspLeuGlyLeuAspLysArgThrI 187

542 TACATTGAAGCACTCCGCGACTCCCTTTTTAAAAGTGGAGCCTTTTCTTC 591
        ||||||| |   :::::: ||       :::::::
187 leGluHisTyrValAspLeuGlnTyrValPro................. 197

592 CTGGACACTATGAAGTTTTGGATTTAAAGCAAATGGCAAAGTTGCATCC 641
        ||||||               |||
198 ...................GluProAspThrLeuHisAlaGl 205

642 GTGGAAATGGTTAAATATCATCACTGTCGGGATGAACCCCTGCACGCCCT 691
        :::::::::||||:                  |||
205 nIleSerArgLeuGluSerGlyCysThrAlaThrValArgProGlyGlyL 222

692 CTATGACAATGTGGGAGAAA..............CTCTTTCCAGGTTTTG 726
        :::::::::::::::                :::  ||||||  :
222 ysLeuGluGlnLysArgTyrPheLysProGlnPheProValGlnLysVal 238

727 AGATAGAA.......................................ACT 737
        |||
238 ys..............................................ACT 737

738 GTGAAGAACAACCTTCAGGATCCTTTTTAAT.................. 767
        ||||||     |||:::    :::::::
239 ValLysGlyLysGluGlnAspLeuPheAspArgIleAlaGlnValLeuGl 255

768 .AATGCTGTAAAGAAAACGTTTGATGACAGACAGAAGGATTGGCTGCCTTT 816
        ::::|||:::||: ::||| ::::|| :||| :::||
255 uAspSerValGluLysHisMetArgAlaAlaAspValThrValGlySerPheL 272

817 TATCAGGGGGCTTGGGACTCCAGCTTGTTGCTGCCACTCTGTTGAAGCAG 866
        |||||||| ||||| ||||||| ||||||||
272 euSerGlyGlyIleAspSerThrAlaIleAla.ProLeuAla........ 285

867 CTGAAAGAAGCCCAAGTACAGTATCCTCTCCAGACACATTTGCAATTGGCAT 916
        :::||||||| ||||||||||  |||  ||||||||||
286 ...................LysArgHis.................... 288

289 ........AsnProAspLeuLeu 293
         :|||||||||||
917 GGGAGAGAMAGCCCCGATTTTACTG 939
```

```
q_name: gb_est2:BF622268

seq_documentation_block:
LOCUS       BF622268              854 bp     mRNA    linear  EST 17-OCT-2001
DEFINITION  HVSMEa0002G16f Hordeum vulgare seedling shoot EST library
            HVcDNA0001 (Cold stress) Hordeum vulgare cDNA clone HVSMEa0002G16f,
            mRNA sequence.
ACCESSION   BF622268
VERSION     BF622268.2  GI:13080047
KEYWORDS    EST.
SOURCE      barley.
ORGANISM    Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 854)
AUTHORS     Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Oates,R., Chol,D.W.
            , Fenton,R.D. and Main,D.
TITLE       Development of a genetically and physically anchored EST resource
            for barley genomics: Morex cold-stressed seedling shoot cDNA
            library
JOURNAL     Unpublished (2001)
COMMENT     On Dec 18, 2000 this sequence version replaced gi:11886002.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: AATTAACCTCACTAAAGGG
            Total hq bases - 574
            High quality sequence stop: 841.
FEATURES             Location/Qualifiers
     source          1..854
                     /organism="Hordeum vulgare"
                     /cultivar="Morex"
                     /db_xref="taxon:4513"
                     /clone="HVSMEa0002G16f"
                     /clone_lib="Hordeum vulgare seedling shoot EST library
                     HVcDNA0001 (Cold stress)"
                     /tissue_type="Seedling shoot"
                     /lab_host="TJC121"
                     /note="Vector: lambdazAP; Site_1: EcoR1; Site_2: Xhol;
                     Seeds were surface sterilized then germinated under axenic
                     conditions in the dark at room temperature on filter paper
                     with water, nystatin and cefotaxime in covered
                     crystallization dishes. Five-day old seedlings were
                     incubated at 5oC for 2 days. Shoots were then harvested,
                     total RNA was prepared, poly(A) RNA was purified, one
                     primary unamplified cDNA library was made, and 600000 pfu
                     were in vivo excised to give pBluescript SK(-) cDNA
                     phagemids. These steps were performed in the TJ Close
                     laboratory at the University of California, Riverside
                     (Chol, Close, Fenton). Phagemids were plated and picked at
                     the Clemson University Genomics Institute (CUGI) (Begum,
                     Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations
                     , DNA sequencing and sequence analysis were performed at
                     CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main
                     ). The sequence has been trimmed to remove vector sequence
                     and contains a minimum of 100 bases of phred value 20 or
                     above. For more details on library preparation and
                     sequence analysis see
                     http://www.genome.clemson.edu/projects/barley. To order
                     this clone see http://www.genome.clemson.edu/orders Also
                     see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                     Genetically and physically anchored EST resources for
                     barley genomics. Barley Genetics Newsletter 31:29-30.
                     (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT      191 a    232 c    235 g    196 t
ORIGIN
```

```
alignment_scores:
            Quality:      263.00          Length:        279
            Ratio:          1.594          Gaps:           13
Percent Similarity:  59.140     Percent Identity:  30.824

alignment_block:
US-09-786-474-2 x BF622268
Align seg 1/1  to: BF622268   from: 1 to: 854

    1 MetCysGlyLeuGlyIleLeuThrAlaAsnGlyAsnAlaGlu..... 15
      |||||||||.|||.||:::::|||
   57 ATGTGTGGCATACTGGCAGTGCTGGCGTCGCCGATGACGACCCAGGGGAA 106

   16 ....AlaPheValProAlaLeuGluArgAlaLeuProCysMetArgHisA 31
      |||         ||::::|||      ||| ||||||
  107 GAGAGTGCCGGTGCTCGAGCTCCGCGCAGGCTC..........AAGCACC 147

   31 rgGlyProAspAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGly 47
      ||||||||    :::|||      ||||| :||||| ||||||
  148 GCGGCCCCGACTGGAGTGGCATGCACCAGGTCGGTGACTCGTCTACCTCTCC 197

   48 PheAsnArgLeuSerIleIleAspIleAlaHisSerHisGlnProLeuAr 64
      :::|||||||||:|||||||| ||| :::    |||||||
  198 CACCAGCGGCTGCCTCCGATCATCGACCCTGCCTCAGGCGACCAGCCACTT.. 245
```

```
FEATURES             Location/Qualifiers
     source          1..927
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DK008YP10"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT        257 a      178 c      213 g      277 t      2 others
ORIGIN

alignment_scores:
         Quality:   259.00              Length:    316
         Ratio:       1.515             Gaps:       13
     Percent Similarity:  54.114    Percent Identity:  26.266

alignment_block:
US-09-786-474-2 x AL555468

Align seg 1/1  to: AL555468  from: 1  to: 927


  2 CysGlyLeuLeuGlyIleLeuThrLeuAsnGlyAsnAlaGlyAlaAlaPheVa 18
 90 TGCACCATGTGCAGCCATTGGGCGCCGTCGTTTGCAGTGATGATTGCCTT.. 137

 18 lProAlaGluArgAlaLeuProCysMet................ArgH 30
138 ...............TCTGTTCAGTGTCTGAGTGCTATGAAGAGATTGCAC 171

 30 isArgGlyProAspAspArgTrpHisAsp............Ala 42
172 ACAGAGGTCCA...GATGCAATTCCGTTTGAGAATGTCAATGGATGGATACACC 218

 43 AspAlaAlaPheGlyPheAsnArgLeuSerIleIleAspIleAlaHisSe 59
219 AACTGCTGCTTTGGATTTCATCGGTTGGCCGGTAGTTGACCCGCTGTTTGG 268

 59 rHisGlnProLeuArg....................TrpGlyProAlaA 69
269 AATGCAGCCATTCCAGTCGAAGAAAATATCCCGTATTTGTGG.......... 308

 69 spGluProAspArgTyrAlaMetThrPheAsnGlyGluIleTyrAsnTyr 85
309 ...............CCTCGTTACAATGGTGAAATCTACAACCAT 338

 86 ValGluLeuArgLysGluAspAspLeuSerLeuGlyTyrThrPheAsnThrSe 102
339 AAGAAGATGCAAACAGCATTTGAA..........TTTGAATACCAGACCAA 379

102 rGlyAspGlyGluProIleValValGlyPheHisTrpGly...GluS 118
380 AGTGGATGGTGAGATAATCCTTCATCTTATGCACAAGGAGGAATTGAGC 429

118 erValValGluHisLeuArgGlyMetPheGlyIleAlaIleTrpAspThr 134
430 AAACAATTGTATGTTGGAGTGTGTTTCCATTTGTTTTGTTTTACTGGATACT 479

135 LysGluLysSerLeuPheLeuAlaArgAspGlnPheGlyIleLysProLe 151
480 GCCAATAAGAAAGTGTTCCTGGGTAGAGATACATATGAGTCAGACCTTT 529

151 uPheTyrAlaThrThrGluHisGly...ThrValPheSerSerGluLysL 167
```

```
 64 gTrpGlyProAlaAspGluProArgTyrAlaMetThrPheAsnGlyG 81
246 ..........TACAACGAGGACAGCAAGTCCATCGTCCTCACTGTGAATGGAG 285

 81 luIleTyrAsnTyrValGluLeuArgLysGluAspAspLeuSerAspLeuGlyTyr 97
286 AGATCTACAACCATGAACAGCTCCGGGCGCCAGCTCTCC......TCCCAC 329

 98 ThrPheAsnThrSerGlyAspGlyGluProIleValValGlyPheHisHi 114
330 ACATTCAGACAGCAGCCAGCCACTCGAGGTCATCGCAGCACCTGTACGAGGA 379

114 sTrpGlyGluSerValValGluHisLeuArgGlyMetPheGlyIleAlaI 131
380 GCATGGGGAGAATTCATCGACACTGTGGATGGTGTCTTCTCCTTCGTCT 429

131 leTrpAspThrLysGluLysSerLeuPheLeuAlaArgAspGlnPheGly 147
430 TGCTCGACACGCGACAACAGCTTCATTGCTGCCGTGATGCCCATTGGT 479

148 IleLysProLeuPheTyrAlaThrThrGluHisGlyThrValPhe...Se 163
480 GTCACACCCATGTATGTTGGCTGGGGAATTGATGGGTCCGGTGTGGATATC 529

163 rSerGluLysLysThrLeuThrIleLeuGluMetAlaGluGluMetAsnLeuAspL 180
530 ATCAGAGATGAAG............................... 542

180 euGlyLeuAspLysGlyLysThrIleGluHisTyrArgThrValAspLeuGlnTyrVal 196
543 ..GGCCTGAATGAT...GATTGTGAGCACTTT.......GAGATCTTT 578

197 ProGluProAspThrLysGluHisAlaGlnIleSerArgLeuGluSerGlyCy 213
579 CCA...CCTGGTCATCTCACTCCAGC.................... 602

213 sThrAlaThrValArgProGlyGlyLysLeuGluGlnLysArgTyrPheL 230
603 ...............AAGGAGGAGGC......TTCAAGAGATGGTACA 630

230 ysProGlnPheProValGlnLysLysValValLysGlyLysGluGlnAsnAspLeu 246
631 ACCCACCTTGG...GTCTCGGAGGTCATTCCTTCAGTGCCATATGATCCA 677

247 PheAspArgIleAlaGlnValAlaLeuGluAspSerSerValGluLysHisMetAr 263
678 CTT...GCTCTCAGGAAAGCTTTCGAATAGGCTGTCACCAAGAGGCTTAT 724

263 gAlaAspValThrValGlySerPheLeuSerGlyGly 275
725 GACAGACGTTCCATTCGGTGTTCTACTCTTTGGGTGGG 761
```

_name: gb_est1:AL555468

```
seq_documentation_block:
LOCUS       AL555468                 927 bp    mRNA    linear   EST 16-FEB-2001
DEFINITION  AL555468 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DK008YP10 5
            prime, mRNA sequence.
ACCESSION   AL555468
VERSION     AL555468.1  GI:12897229
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 927)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
```

University, The Netherlands). The cSTA libraries will attempt to capture the induction and initiation/initial growth of the tuber in an in vitro system as described in Bachem et al. (Plant Journal, 1996). Small microtubers develop from axillary buds attached to stem explants when placed on a high sucrose medium (10%). Visible morphological changes occur synchronously at day five in the axillary buds. The first library, cSTA (1-20) consists of axillary buds harvested on days 1-3. This targets those genes involved in induction of the microtubers. The following libraries, cSTA (21-40) and cSTA (41-60), capture genes involved in tuber initiation and outgrowth. This library is noted as P3 in Tanksley lab notebooks."

```
BASE COUNT      191 a     133 c     168 g     215 t
ORIGIN

alignment_scores:
         Quality:    255.50         Length:    175
           Ratio:     2.147           Gaps:      3
Percent Similarity:  68.000    Percent Identity:  35.429

alignment_block:
US-09-786-474-2 x BI179230

Align seg 1/1   to: BI179230   from: 1   to: 707

  1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPh  17
    ||||||||||:||||::||    ||||:    ::|:    :|||:|||
 11 ATGTGTGGAATACTTGCAATTTCGGTTGCACTGATAATTCTCATGCCAA   60

 17 eValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA  34
    ::|||||||||    ::|    :||||||    :||    ||||:|||
 61 GGCGTTCAAGAGATCATCGAACTATCAAGAAGGTTGCGCCATAGAGGACCTG  110

 34 spAspAlaGlyThrTrpHisAspAlaAlaGlyPheGlyPheAsnAsnArg  50
    ||    |||:|    ::||    |:    ::||:    ::|||
111 ATTGGAGTGGATTGCATAGCCATGAGGACTGTATCTTGCTCATCAACGA  160

 51 LeuSerIleAspIleAlaHisSerHisGlnProLeuArgTrpGlyPr    67
    ||::||||||    :::    :::    ||    ||||||||
161 TTGGCAATAGTAGACCCAACTTCAGGGGATCAGCCGCTG..........  199

 67 oAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyGluIleTyrA  84
    ::    :::||:    ::|:||||    :::    :::|||
200 .TATAATGAGGACAAGACCATTGTTGTGTCGGTAAATGGAGAGATCTACA  248

 84 snTyrValGlyLeuArgLysGluLeuSerAspLeuGlyTyrThrPheAsn 100
    ||:::||||||    ||:::||||||    |||    ||| :::
249 ACCATAAGGAATTACGGGAGAAACTGAAG......TCTCATCAGTTTGA  292

101 ThrSerGlyAspGlyGluProIleValValGlyPheHisHisTrpGlyGl 117
    ||:::|    ||| |||    :::    :::||:|||||
293 ACTGAAAGTGATTGTGAAGTTATTGCCCATCTTTATGAAGAATATGGAGA 342

117 uSerValValGluHisLeuArgGlyMetPheGlyIleAlaIleTrpAspT 134
    |:::    ||||||||||    |||    :|||
343 AAACTTCATTGACATTGGATGGGATTTCTCTTTGTTCTTCTGATA     392

134 hrLysGluLysSerLeuPheLeuAlaArgAspGlnPheGlyIleLysPro 150
    |:::|    |||||||||    |||||:    ||||||||
393 CCCGGGATAAAAGTTTCATTGCTGCTCCGGATGCCATTGGCATTACACCC 442

151 LeuPheTyrAlaThrThrGluHisGlyThrVal...PheSerSerGluLy 166
    |||    :::    ::|||    ||||||    :|||||||||
443 CTTTATATGGGGTGGGGTCTTGATGGCTCCATATGGTTTCCTCAGAGAT  492

166 sLysThrIleLeuGluMetAlaGlu 174
    :||:||    ::|||
493 GAAAGCCTTAAGTGATGATTGTGA 517

seq_name: gb_est2:BI269282
```

```
530 GTTTAAAGCAATGACAGAAGATGGATTTTGGCTGTATGTTCAGAAGCTA 579
167 ysThrIleLeuGluMetAlaGluGluMetAsnLeuAspLeuGlyLeuAsp 183
    ||    :::::|    ::|:|||    |||    ::|
580 AAGGTCTGTTACATTGAAGCACTCCGCGACTCCCTTTTTAAAAGTGGAG 629

184 LysArgThrIleHisTyrValAspLeuGlnTyrValPro.......    197
    :::||||||    |||||    |||
630 CCTTTTCTTCCTCGGACACTATGAAGTTTGGATTTAAAGCCAAATGCAA 679

198 ..........................GluProAspThrL            202
                              ::|    :::|
680 AGTTGCATCCGTGGAAATGGTTAAATATCACTGTCGGGATGAACCCC    729

202 euHisAlaGlnIleSerArgLeuGluSerGlyCysThrAlaThrValArg 218
    ||||||    ::::|:|||||:::|
730 TGCACCCCCTCTATGACAATGTGGGAGAAA..............CTCTTT 764

219 ProGlyGlyLysLeuGlnLysArgTyrPheLysArgProGlnPheProVa 235
    ||||||    :::::|||
765 CCAGGTTTGAGATAGAA.............................    782

235 lGlnLysValValLysGlyLysGluGlnAspPheAspPheArgIleAlaG 252
    ::||||||:    :::    ||||||
783 ......ACTGTGAARAACAACCTCAGGATCCTTTTTAAT..........  815

252 lnValLeuGluGluAspSerValGluLysHisMetArgAlaAspValThrVal 268
       :::::||||||:||    ::||:||    :|:||
816 ..........AATGCTGTAAAGAAACGTTTGATGACAGACAGAAAGATT 854

269 GlySerPheLeuSerGlyGlyIleLeuAspSerThrAlaIleAlaProLeu 284
    ::|||||||||||||    |||||    :::|    |||
855 GGCTGCCCTTTTATCAGGGGGCTTGGACTCCCAGCTCCTTGGTTGCTGCACTC 902

seq_name: gb_est2:BI179230
```

seq_documentation_block:
LOCUS   BI179230   707 bp   mRNA   linear   EST 09-JUL-2001
DEFINITION EST520175 cSTE Solanum tuberosum cDNA clone cSTE17F18 5' sequence, mRNA sequence.
ACCESSION   BI179230
VERSION   BI179230.1 GI:14645041
KEYWORDS   EST.
SOURCE   potato.
ORGANISM   Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE   1 (bases 1 to 707)
AUTHORS   van der Hoeven,R., Bezzerides,J., Bachem,C., Visser,R., Cho,J., Chlemingo,A., Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.
TITLE   Generation of ESTs from in vitro grown microtubers
JOURNAL   Unpublished (2001)
COMMENT   Contact: Cathy Ronning
            The Institute for Genomic Research
            For clone info: please contact Research Genetics, Libraries
            Division tel 1-800-711-6195, email cdna@resgen.com
            Seq primer: M13F-R.
FEATURES   Location/Qualifiers
     source   1..707
                     /organism="Solanum tuberosum"
                     /cultivar="Bintje"
                     /db_xref="taxon:4113"
                     /clone="cSTE17F18"
                     /clone_lib="cSTE"
                     /tissue_type="axillary buds of stem explants; growing sink-tubers"
                     /dev_stage="7, 8 and 10 days"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Tissue supplied by Christian Bachem and Richard Visser (Department of Plant Breeding, Wageningen

seq_documentation_block:
LOCUS   BI269282
DEFINITION   NF006E11R1F1086 Irradiated Medicago truncatula cDNA clone
              NF006E11R 5', mRNA sequence.
ACCESSION   BI269282
VERSION   BI269282.1 GI:14875861
KEYWORDS   EST.
SOURCE   barrel medic.
ORGANISM   Medicago truncatula
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
              Medicago.
REFERENCE   1 (bases 1 to 699)
AUTHORS   Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
              Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
TITLE   Expressed Sequence Tags from the Samuel Roberts Noble Foundation
              Medicago truncatula irradiated library
JOURNAL   Unpublished (2001)
COMMENT   Contact: May GD
              Plant Biology Division
              The Samuel Roberts Noble Foundation
              2510 Sam Noble Parkway, Ardmore, OK 73402, USA
              Tel: 580 221 7391
              Fax: 580 221 7380
              Email: gdmay@noble.org
              Insert Length: 699   Std Error: 0.00
              Plate: 006 row: E column: 11
              Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES   Location/Qualifiers
     source   1..699
                     /organism="Medicago truncatula"
                     /db_xref="taxon:3880"
                     /clone="NF006E11R"
                     /clone_lib="Irradiated"
                     /tissue_type="seedlings"
                     /dev_stage="seedling"
                     /note="Vector: Lambda Zap; Seedlings were exposed either
                     to 100 gy gamma or 0.5, 1, 5, or 10 kJ/m2 UV irradiation.
                     Gamma-irradiated samples were harvested at 6, 12, 24 and
                     48 hours after treatment. UV-irradiated samples were
                     harvested 24 hours post-treatment. cDNA was prepared from
                     polyA+ enriched, pooled samples of equivalent amounts of
                     total RNA from each sample. The cDNA was directionally
                     ligated into the Uni-Zap XR vector (Stratagene) and
                     packaged using the Gigapack III Gold packaging extracts.
                     Phagemids containing cDNA inserts were in vivo excised
                     from the recombinant Uni-ZAP XR vector using ExAssist
                     helper phage and the E. coli strain XLl-Blue MRF'
                     (Stratagene). Excised plasmids were plated using SOLR
                     cells."

COUNT   156 a   173 g   190 c   178 t   2 others
IN

alignment_scores:
     Quality: 254.00   Length: 267
     Ratio: 1.618   Gaps: 8
Percent Similarity: 58.801   Percent Identity: 27.341

alignment_block:
US-09-786-474-2 x BI269282/rev

Align seg 1/1 to reverse of: BI269282 from: 1 to: 699

74 TyrAlaMetThrPheAsnGlyGlyIleTyrAsnTyrValGluLeuArgLy 90
    ::::::::::::::: |||||||| ||||||| ||||||
684 CACGTACTGGCCGGTAAACGGTGAAATNTACAACCACCAGGCATTGCGCGC 635

90 sGluLeuSerAspLeuGlyTyrThrPheAsnThrSerGlyAspGlyGluP 107
    ||| :::||| ::||| ||::::|||:::::|||| ||| |||
634 CGAATATGGCGAT...CGTTACCAGTTCCAGACCGGGTCTGACTGTGAAG 588

rolleValValGlyPheHisHisTrpGlyGluSerValValGluHisLeu 123 107
||| ||||| :::::::| || ::::: || |||
TGATCCTCCGCCGTGTATCAGGAAAAAGGGCCGGAATTCTCCGACGACTTG 538 587

ArgGlyMetPheGlyIleAlaIleTrpAspThrLysGluLysSerLeuPh 140 124
:::||||||||| |||||||||||||:::
CAGGGCATGTTTGCCTTTGCACTGTACGACGCGAAAAAGATGCTACCT 488 537

eLeuAlaArgAspGlnPheGlyIleLysProLeuPheTyrAlaThrThrG 157 140
:::::::|||:||||::::|||||||||||
GATTGGTCCGACCATCTGGGATCATCCCACTGTATATGGGTATGACG 438 487

luHisGlyThrValPhe...SerSerGluLysLysThrIleLeuGluMet 172 157
||||||||| :::||| :::|||||||
AACACGGTCAGCTGTATGTGGCCTCAGAAATGAAAGCGCCTGGTGCCAGTT 388 437

AlaGluGluMetAsnLeuAspLeuGlyLeuAspLysArgThrIleGluHi 189 173
|||||||| CGCACGATTAAAGA 371
:::
ACGGCGAAAATCCGTTCT 322 387
TGC........

sTyrValAspLeuGlnTyrValProGluProAspThrLeuHisAlaGlnI 206 189
::: ||| :::|||
GTTC...........CCGGCGGGAGCTATTTGTGGAGCCAGG 339 370

leSerArgLeuGluSerGlyCysThrAlaThrValArgProGlyGlyLys 222 206
:::::::
ACGGCGAAATCCGTTCT 322 338

LeuGluGlnLysArgTyrPheLysProGlnPheProValGlnLysValVa 239 223
                       ::||
TACTATCATCGCGACTGGTTCGACTACGATGCGGT 287 321

lLysGlyLysGluAspPheAspArgIleAlaGlnAlaLeuGluGluA 256 239
|||| :: ||| :::::::::||||||||
GAAAGATAACGTGACCGACAAA..AACGAGCTGCCTCAGGCACTCGAAG 240 286

spSerValGluLysHisMetArgAlaAspValThrValGlySerPheLeu 272 256
||||||||||::::||| ||| ||:::||||||::|:|
ATTCAGTTAAAGCCATCTGATGTCGATGTGCCTTACGGTGTGCTT 190 239

SerGlyGlyIleAspSerThrAlaIleAlaAlaProLeuAlaLysArgHisAs 289 273
|||||||||| |||| ||| :: :::||| |||::||||
TCTGGTGGTCTGGATTCCTCAATTATTTCCGCTATCACCAAGAAATACGC 140 189

....ProAspL 292 289
                                         |||::||
AGCCCGTGCGGTGGAAGATCAGGAAACGCTCGAAGCCTGGTGGCCCGCAGT 90 139

euLeuThrPheThrGlyPheGluArgGluGlyTyrSerGluValAsp 308 292
||| :::::||||| ||||||| ::::
TACACTCCTTTGCTGTAGGTCTGCCG..........GGTCACCGGAT 52 89

ValAlaAlaGluSerAlaAlaAlaIleGlyAlaAlaGluHisIleValLysIl 325 309
||| :::|| ::||| |||||
CTGAAAGCAGCCCAGGAAGTGGCATCCCAGGCACAGAACATGAATAAAAT 2 51

325 e 325
  | -
  1 C 1

seq_name: gb_est1:AL515757

seq_documentation_block:
LOCUS   AL515757
DEFINITION   AL515757 LTI_NFL011_NBC1 Homo sapiens cDNA clone CS0DA001YC08 5
              prime, mRNA sequence.   1092 bp   mRNA   linear   EST 13-FEB-2001
ACCESSION   AL515757
VERSION   AL515757.1 GI:12779250
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```
REFERENCE   1 (bases 1 to 1092)
AUTHORS     Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES    Location/Qualifiers
source      1..1092
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="CS0DA001YC08"
            /clone_lib="LTI_NFL011_NBC1"
            /sex="male"
            /tissue_type="neuroblastoma cells"
            /lab_host="DH10B"
            /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
            was primed with a NotI-oligo(dT) primer. Five prime end
            enriched, double-stranded cDNA was digested with Not I and
            cloned into the Not I and Eco RV sites of the pCMVSPORT 6
            vector. Library is not normalized, but is the control for
            the normalized libraries. Library was constructed by Life
            Technologies. Contact : Feng Liang Life Technologies, a
            division of Invitrogen 9800 Medical Center Drive Rockville
            , Maryland 20850, USA Fax : (1) 301 610 8371 Email :
            fliang@lifetech.com URL :
            http://fulllength.invitrogen.com"
BASE COUNT   309 a    206 c    251 g    318 t      8 others
ORIGIN
```

alignment_scores:
    Quality: 253.50   Length: 363
    Ratio: 1.268   Gaps: 14
    Percent Similarity: 55.096   Percent Identity: 26.722

alignment_block:
US-09-786-474-2 x AL515757

Align seg 1/1 to: AL515757   from: 1 to: 1092

(sequence alignment data)

seq_name: gb_est2:BM457834

```
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Lou Staudt
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM12131 row: 1 column: 20
             High quality sequence stop: 687.
FEATURES     Location/Qualifiers
     source  1..870
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="IMAGE:5499043"
             /clone_lib="NIH_MGC_85"
             /tissue_type="lymphoma, cell line"
             /lab_host="DH10B (phage-resistant)"
             /note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
             Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
             Average insert size 1.867 kb. Library enriched for
             full-length clones and constructed by Life Technologies.
             Note: this is a NIH_MGC Library."
BASE COUNT   237 a  162 c  211 g  260 t
ORIGIN
```

alignment_scores:
```
Quality:              253.00          Length:             314
Ratio:                1.497           Gaps:               13
Percent Similarity:   53.822          Percent Identity:   26.115
```

alignment_block:
US-09-786-474-2 x BM457834

Align seg 1/1 to: BM457834 from: 1 to: 870

```
  2 CysGlyLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPheVa 18
    ||| .:: ||||||  ||: ||||||:::::::::::::
 22 TGCACCATGTGTGCCATTTGGCCCTGTTCGGCAGTGATGATTGCCTT.. 69

 18 lProAlaLeuGluArgAlaLeuProCysMet.............ArgH 30
    ::: .|||:::
 70 ............TCTGTTCAGTGTCTGAGTGCTATGAAGATTGCAC 103

 30 isArgGlyProAspAspAlaGlyThrThrProHisAsp..........Ala 42
    ||||||||||  |||::
 30 isArgGlyProAspAspAlaGlyThrThrProHisAsp :::
104 ACAGAGGTCCA...GATGCATTCCGTTTGAGAATGTCAATGGATACACC 150

 43 AspAlaAlaPheGlyPheAsnArgLeuSerIleIleAspIleAlaHisSe 59
    :::::|||||||:||:||||:||
151 AACTGCTGCTTTGGATTTCACCGGTTGCGGTAGTTGACCCCGCTGTTGG 200

 59 rHisGlnProLeuArg..................TrpProGlyProAlaA 69
    |||:||:||
201 AATGCAGCCAATTCGAGTGGAGAAATATCCGATATTGTGG........ 240

 69 spGluProAspArgTyrAlaMetThrPheAsnGlyLeuGlyLleTyrAsnTyr 85
    :: ||||||||||||
241 ............CTCTGTTACAATGGTGAAATCTACAACCAT 270

 86 ValGluLeuArgLysGluLeuSerAspLeuGlyTyrThrProAsnThrSe 102
    ::: ||||:||||||
271 AAGAAGATGCAACAGCATTTGAA......TTTGAATACCAGACCAA 311

102 rGlyAspGlyProIleValValGlyPheHisHisTrpGly...GluS 118
    : ||||| ||:: |||
312 AGTGGATGGTGAGATAATCCTTCATCTTATGACAAAGGAGGAATTGAGC 361

118 erValValGluHisLeuArgGlyMetPheGlyIleAlaIleTrpAspThr 134
    ::::::: ||| |||||::|||:::::::::
```

```
362 AAACAATTGTATGTTGGATGGTGTGTTGCATTGTTTGTTACTGGATACT 411

135 LysGluLysSerLeuPheLeuAlaArgAspGlnPheGlyLleLysProLe 151
    :::||||:::|||||| :::|||::::|||||
412 GCCAATAAGAAAGTGTTCCTGGCTAGAGATACATATGGAGTCAGACCTTT 461

151 uPheTyrAlaThrThrGluHisGly...ThrValPheSerGluLysL 167
    |||| ||| |||||| ||||| ||||||
462 GTTTAAAGCAATGACAGAAGATGGATTTTGGCTGTATGTTCAGAAGCTA 511

167 ysThrIleLeuGluMetAlaGluGlumetAsnLeuAspLeuGlyLeuAsp 183
    ||| :.::::  |: :::|| :::
512 AAGGTCTTGTTACATTGAAGCACTCCGCGACTCCCTTTTAAAAGTGGAG 561

184 LysArgThrIleGluHisTyrValAspLeuGlnTyrValPro....... 197
    ||||||
562 CCTTTTCTTCCTGGACACTATGAAGTTTGGATTAAAGCCAAATGCAA 611

198 ..............................GluProAspThrL 202
    ::: :::
612 AGTTGCATCCGTGGAAATGGTAAATATCATCACTGTCGGGATGAACCCC 661

202 euHisAlaGlnIleSerArgLeuGluSerGlyCysThrAlaThrValArg 218
    |||||| |:: :||| :::
662 TGCACGCCCTCTATGACAATGTGGAGAAA.............CTCTTT 696

219 ProGlyGlyLysLeuGluGlnLysArgTyrPheLysProGlnPheProVa 235
    |||||  :::::::||
697 CCAGGTTTTGAGATAGAA................................ 714

235 lGlnLysValValLysGlyLysGluGlnAspLeuPheAspArgIleAlaG 252
    :::||||||:::  ||:: ||||| |||||
715 .........ACTGTGAAGAACAACCTCAGGATCCTTTTT.......... 744

252 lnValLeuGluAspSerValGluLysHisMetArgAlaAspValThrVal 268
    ||| :::  || |||::
745 ........TATAATGCTGTAAAGAAACGTTTGATGACAGACAGAAGGATT 786

269 GlySerPheLeuSerGlyGlyIleAspSerThrAlaIleAla 282
    .::|||||||||||||:||::||| :::||
787 GGCTGCCTTTTATCAGGGGGCTTGGACTCCAGCTTGGTTGCT 828
```

```
seq_name: gb_est2:BE325921

seq_documentation_block:
LOCUS       BE325921                   675 bp    mRNA        linear   EST 14-JUL-2000
DEFINITION  NF084F12STIF1096 Developing stem Medicago truncatula cDNA clone
            NF084F12ST 5', mRNA sequence.
ACCESSION   BE325921
VERSION     BE325921.1   GI:9199698
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 675)
  AUTHORS   He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell
            ,C.J., Flores,H.R., Inman,J.T., Weller,J.W., May,G.D. and Dixon
            ,R.A.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula stem library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Dixon RA
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7302
            Fax: 580 221 7380
            Email: radixon@noble.org
            Medicago Genome Initiative accession: MGI:S:26492
            Insert Length: 675    Std Error: 0.00
```

FEATURES
    source   Plate: 084 row: F column: 12
                    Seq primer: TCACACAGAAAACAGCTATGAC.
                    Location/Qualifiers
                    1..675
                    /organism="Medicago truncatula"
                    /db_xref="taxon:3880"
                    /clone="NF084F12ST"
                    /clone_lib="Developing stem"
                    /tissue_type="stem"
                    /dev_stage="Pooled developmental"
                    /note="Vector: Lambda Zap; Contains a mixture of
                    internodal stem segments"

BASE COUNT   180 a   129 c   161 g   205 t
ORIGIN

gnment_scores:
    Quality: 249.00   Length: 265
    Ratio: 1.649   Gaps: 7
Percent Similarity: 56.981   Percent Identity: 27.547

alignment_block:
US-09-786-474-2 x BE325921   .

Align seg 1/1 to: BE325921 from: 1 to: 675

3   GlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaAlaGluAlaPheValPr 19
    |||::|||||::|| |||
2   GGAATCTTAGCAGTGTTGGGTGTGTCGACAATTCCAGGCCAAACGCGC 51

19 oAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProAspAspA 36
    ::: :::   :::||||||||||||||||
52 TCGCCATCATCGAATTGTCTCCGTAGGTTGCGACATAGAGGTCCTGATTGGA 101

36 laGlyThrTrpHisAspAlaAspAlaAlaPheGlyPheAsnArgLeuSer 52
    :|||| ::|:|| |||||::::
102 GTGGTTTGCATTGTCATCAAGATTGTACCTCCGCTCATCAACGTCTTGCT 151

53 IleIleAspIleAlaHisSerHisGlnProLeuArgTrpGlyProAlaAs 69
    ||||||||||
152 ATTGTGGACCCTACTTCCGGAGATCAACCTCTT..........TACAA 189

69 pGluProAspArgTyrAlaMetThrPheAsnGlyGluIleTyrAsnTyrV 86
    ::||| |||||||||||||||:|||::
190 CTATGACAACACCGGTTATGTCACCCGTTAATGGGGAGATATACAACCACA 239

86 alGlyLeuArgLysGluSerAspLeuGlyTyrTyrPheAsnThrSer 102
    ::|||||||::::: |||||||:
240 AGAAATTGAGGCAGAAACTGAAA.......TCTCATCAATTTCGAACTGGT 283

103 GlyAspGlyGlyValProIleValValGlyPheHisHisTrpGlyGluSerVa 119
    ::||| ||| ||:||
284 AGTGACTGTGAAGTCATTGCTCATCTTTATGAAGAATATGGTGAAGAGTT 333

119 lValGluHisLeuArgGlyMetPheGlyIleAlaIleTrpAspThrLysG 136
    :::|| ||| ||||||||||
334 TGTTGATATGCTGGATGGTTCTCATTTGTCCTTCTTGTCGTATACTTCGAG 383

136 luLysSerLeuPheLeuAlaArgAspGlnPheGlyIleLeuLysProLeuPhe 152
    ::|||||:::: |||||||| ||||:::
384 ATAAAAGTTTCATTGCTGCGTGATGCTATTGGCATTACCCCTCTTTAC 433

153 TyrAlaThrThrGluHisGlyThrVal...PheSerSerGluLysLysTh 168
    ::: |||||||||| ||||:::
434 TTGGGTTGGGCCATGATCAATATGGTTTGCGTCTGAAATGAAAGC 483

168 rIleLeuGluMetAlaGluGluMetAsnLeuAspLeuGlyLeuAspLysA 185
    ::::::::: :::|||
484 TCTGATTGACCGATTGTGAG............................. 502

185 rgThrIleGluHisTyrValAspLeuGlnTyrValProGluProAspThr 201
    ||| |||
503 ........................CAATTCATATCTTTTCCT...... 520

<hr/>

202 LeuHisAlaGlnIleSerArgLeuGluSerGlyCysThrAlaThrValAr 218
    :|| ::   :::
521 .................................CCAGGGCATATATTCCAGCAA 543

218 gProGlyGlyLysLeuGluGlnLysArgTyrPheLysProGlnPheProV 235
    : ||| ::: :::
544 ACAGGGAGGA..........TTAAGAAGGTGGTATAATCCACCATGGTTCA 584

235 alGlnLysValValLysGlyLysGluGlnAsnLeuPheAspArgIleAla 251
    :::||||:|:   ::: :::   :::
585 CAGAGAAAATTCCATCAACACGCTATGATCCAACGGTT......TTGCGT 628

252 GlnValLeuGluAspSerValGlyLysHisMetArgAlaAspVal 266
    :::::||| :::||| :::|||||||
629 GAAGCCTTTGAGAGGGCTGTAGTTAAGAAAATGATGACTGATGTA 673

seq_name: gb_est2:BF618750

```
preparations, DNA sequencing and sequence analysis were
performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates
, Rambo, Main). The sequence has been trimmed to remove
vector sequence and contains a minimum of 100 bases of
phred value 20 or above. For more details on library
preparation and sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders Also
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

BASE COUNT      180 a      207 c      227 g      174 t
ORIGIN

alignment_scores:
      Quality:    249.00          Length:    264
        Ratio:      1.638           Gaps:       8
Percent Similarity:  57.576     Percent Identity:  28.030

alignment_block:
US-09-786-474-2 x BF618750     ..

Align seg 1/1  to: BF618750  from: 1  to: 788

   1 MetCysGlyLeuGlyIleLeuThrAlaAlaAsnGlyAsnAlaGlu...... 15
     |||||||||||||:::|||::::::::                :::::::
  44 ATGTGTGGCGTGCTGGCAGTGCTGGGCTGCGCCGATGACACCCAGGGGAA 93

  16 ...AlaPheValProAlaLeuGluArgAlaLeuProCysMetArgHisA 31
     ::: |||  |||:|||  |                       :::|||||
  94 GAGAGTGCGGGGTGCTCGAGCCTCCGGCCAGGCTC........AAGCACC 134

  31 rgGlyProAspAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGly 47
     :::|||               ||| :::|||||||
 135 GCGGCCCGACTGGAGTGGCATGCACCAGGTCGGTGACTGCTACCTCTCC 184

  48 PheAsnArgLeuSerIleIleAspIleAlaHisSerHisGlnProLeuAr 64
     :::||||||||:::||:||  |||||    ||||:|||||||||
 185 CACCAGCGCCTGCCCATCATGGACCCTGCCTCAGGCGACCAGCCACTT.. 232

  64 gTrpGlyProAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyG 81
     :::|||               ||||||||||:||:||||:::||||
 233 ........TACAACGAGGACAAGTCCATCGTCGTCACTGTGAATGGAG 272

  81 luIleTyrAsnTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyr 97
     ||||||||:::||:|||||:||  :::||||| ::::||||
 273 AGATCATACAAACCATGAACAGCTCCGGGCGCCAGCTCTCC.....TCCCAC 316

  98 ThrPheAsnThrSerGlyAspGlyGluProIleValGlyPheHisHi 114
     ||||||:::||||| :::  :||::::||
 317 ACATTCAGGACAGCAGCCACTGCGAGTCATCGCAGCCACCTGTACGAGGA 366

 114 sTrpGlyLeuSerValValGluHisLeuArgGlyMetPheGlyIleAlaI 131
     :::|||      |||               ||||||||:||||
 367 GCATGGGGAGGAGAATTCATCGACACATGCTGGATGGTGTCTTCCTTCGTCT 416

 131 leTrpAspThrLysGluLysSerLeuPheLeuAlaLeuArgAspGlnPheGly 147
     |||||||:::|||:||||||   ||||:||||||  |||||||::  ||||||
 417 TGCTCGACACGCGACAACAGCTTCATTGCTGCCCCGTGATGCCATTGGT 466

 148 IleLysProLeuPheTyrAlaThrThrGluHisGlyThrValPhe...Se 163
     :::|||:||||| ::: |||| ||| |||||||:||||||         |
 467 GTCACACCCATGTATGTTGGCTGGGGAATTGATGGGTCCGTGTGGATATC 516

 163 rSerGluLysThrIleLeuGluMetAlaGluGluMetAsnLeuAspL 180
     |||||||||  |||||||||||||  |||:||::: :||::  ::::
 517 ATCAGAGATGAGGGGCCTGATTGAT................ 541

 180 euGlyLeuAspLysArgThrIleGluHisTyrValAspLeuGlnTyrVal 196
     :::|||||| |||:|||||||:|||||||||||||||||||||||||||
 541 ........... 541
```

```
 197 ProGluProAspThrLeuHisAlaGlnIleSerArgLeuGluSerGlyCy 213
     |||                               :: :||||
 542 ........GATTGTAAGCACTTGAGATTCTT..........CCACCCTG 573

 213 sThrAlaThrValArgProGlyLysLeuGluGlnLysArgTyrPheL 230
     :::  :::  |||||||:||||| |||                 |||
 574 GTCATCTTACTCCAGCCAGGAGGAGGCCTTCAAGAAGAAGGA...... 616

 230 ysProGlnPheProValGlnLysValValLysGlyGluGlnAspLeu 246
     |||                               :: :::|||
 617 ..........CAACCCACCTTGGGTCGTCGAGGCATTCCTTG 649

 247 PheAspArgIleAlaGlnValLeuGluGluValGluLys 260
     :::  :::  ::||||||               |||||||
 650 AAGGCCAATGTGATACACTTGCTTTCAGGAAAGCTTGAGAAA 691


seq_name: gb_est1:AV925031

seq_documentation_block:
LOCUS       AV925031               571 bp    mRNA        linear   EST 18-JAN-2002
DEFINITION  AV925031 K. Sato unpublished cDNA library, cv. Haruna Nijo second
            leaf stage seedling leaves Hordeum vulgare subsp. vulgare cDNA
            clone basd21a10 5', mRNA sequence.
ACCESSION   AV925031
VERSION     AV925031.1  GI:18220810
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare.
ORGANISM    Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 571)
  AUTHORS   Sato,K., Saisho,D. and Takeda,K.
  TITLE     Barley EST sequencing project in NIG and Okayama Univ
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..571
                     /organism="Hordeum vulgare subsp. vulgare"
                     /cultivar="Haruna Nijo"
                     /db_xref="taxon:112509"
                     /clone="basd21a10"
                     /clone_lib="K. Sato unpublished cDNA library, cv. Haruna
                     Nijo second leaf stage seedling leaves"
                     /tissue_type="seedling leaves"
                     /dev_stage="second leaf stage"
BASE COUNT      123 a      162 c      164 g      122 t
ORIGIN

alignment_scores:
      Quality:    243.50          Length:    182
        Ratio:      1.964           Gaps:       5
Percent Similarity:  68.132     Percent Identity:  35.165

alignment_block:
US-09-786-474-2 x AV925031     ..

Align seg 1/1  to: AV925031  from: 1  to: 571

   1 MetCysGlyLeuGlyIleLeuLeuThrAlaAlaAsnGlyAsnAlaGlu...... 15
     |||||||||||||                                :::|||||
   1 ATGTGTGGCATACTGGCAGTGCTGGGCTGCGCCGATGACACCCAGGGGAA 100

  16 ....AlaPheValProAlaLeuGluArgAlaLeuProCysMetArgHisA 31
     :::                                    :::::|||||
  51 ATGTGTGGCATACTGGCAGTGCTGGGCTGCGCCGATGACACCCAGGGGAA 100

  16 ....AlaPheValProAlaLeuGluArgAlaLeuProCysMetArgHisA 31
     :::                                    :::::|||||
```

```
101 GAGAGTGCGCCGTGCTCGAGCTCTCGCGCAGGCTC.........AAGCACC 141

 31 rgGlyProAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGly  47
    ||||||||| |||       :::  :||  :::::|||||
142 GCGGCCCGACTGGAGTGGCATGACCAGGTCGGTGACCTGTACCTCTCC 191

 48 PheAsnArgLeuSerIleIleAspIleAlaHisSerHisGlnProLeuAr  64
    :::||||||:::|||||||| ||| :::  ||||||||
192 CACCAGGCGCTCGCCATCATCGACCCTGCCTCAGGCGACCAGCCACTT.. 239

 64 gTrpGlyProAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyG  81
    ::||||      |||||||| |||||||
240 ........TACAACGAGGACAAGTCCATCGTCGTCACTGTGAATGGAG  279

 81 luIleTyrAsnTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyr  97
    ||||||||:: ::|||||| :::|||||  :::||||
280 AGATCTACAACCATGAACAGCTCCGGGCGCCAGCCTCCC.....TCCCAC 323

 98 ThrPheAsnThrSerGlyAspGlyGluProIleValValGlyPheHisHi  114
    ||||||:::::::|||  ||::  |||||    :::|||||
324 ACATTCAGGACAGGCAGCGACTGCCAGGTCATCGCACCACCTGTACGAGGA 373

114 sTrpGlyGluSerValValGluHisLeuArgGlyMetPheGlyIleAlaI  131
    |||||||  :::::::   |||::|||:  |||||||||
374 GCATGGGGAGAATTCATCGACATGCTGGATGGTGTCTTCTCCCTTCGTCT 423

131 leTrpAspThrLysGluLysSerLeuPheLeuAlaArgAspAspGlnPheGly  147
    :: ||||||:::||:||||  |||||||| ::||||
424 TGCTCGACACGCGACACAGCTTCATTGCTGCCCGTGATGCCATTGGT 473

148 IleLysProLeuPheTyrAlaThrThrGluHisGlyThrValPhe...Se  163
    ::: |||||||:: :::|| ||::::|:: :||||:::|| |
474 GTCACACCCCTGTATGTTGGCTGGGGAATTGATGGGTCGGTGTGGATATC 523

163 rSerGluLysLysThrIleLeuGluMetAlaGluGluMetAsnLeu  178
    :::  ||| |||   ::|||||::::::||||
524 ATCAGAGATGAAGGGCCTGAATGATGATTGTGAGCACTTTGAGATC 569

seq_name: gb_est2:BE405536

seq_documentation_block:
LOCUS       BE405536          604 bp    mRNA      linear   EST 21-JUL-2000
DEFINITION  WHE1212_E01_J02ZS Wheat etiolated seedling root cDNA library
            Triticum aestivum cDNA clone WHE1212_E01_J02, mRNA sequence.
ACCESSION   BE405536
VERSION     BE405536.1  GI:9365004
KEYWORDS    EST.
SOURCE      bread wheat.
ORGANISM    Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 604)
AUTHORS     Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
            ,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
            Seaton,C.L. and Tong,J.C.
TITLE       The structure and function of the expressed portion of the wheat
            genomes
JOURNAL     Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer,
FEATURES             Location/Qualifiers
     source          1..604
                     /organism="Triticum aestivum"
                     /cultivar="Chinese Spring"
```

```
                     /db_xref="taxon:4565"
                     /clone="WHE1212_E01_J02"
                     /clone_lib="Wheat etiolated seedling root cDNA library"
                     /tissue_type="Root"
                     /dev_stage="Five day old etiolated seedling"
                     /lab_host="E. coli SOLR"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid:
                     Site_1: EcoRI; Site_2: XhoI; Seeds were surface-sterilized
                     germinated and grown aseptically in the dark at room
                     temperature on filter paper with water, nystatin and
                     cefotaxime in covered crystallization dishes. Roots were
                     harvested.  The tissue, total RNA, and poly(A) RNA were
                     prepared, a cDNA library was made, and the cDNA clones
                     were in vivo excised to give pBluescript phagemids in the
                     TJ Close lab (Choi, Close, Fenton) at the University of
                     California, Riverside. Plasmid DNA preparations and DNA
                     sequencing were performed in the OD Anderson lab (all
                     other authors)."
BASE COUNT      131 a     179 c     169 g     125 t
ORIGIN

alignment_scores:
          Quality:    243.50              Length:        182
          Ratio:      1.964               Gaps:          5
Percent Similarity:  68.132     Percent Identity:  35.165

alignment_block:
US-09-786-474-2 x BE405536

Align seq 1/1   to: BE405536    from: 1   to: 604

  1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGlu......  15
    |||||||  :::::::   ||:|||:  |||||:::|||
 22 ATGTGCGGCATACTGGGCGTGCTGGGCTGCGCTGATGACACCCAAGGAA  71

 16 ....AlaPheValProAlaLeuGluArgAlaLeuProCysMetArgHisA  31
    ||   |||||||:::||| ||  :::|||
 72 GAGAGTGCGCCGTGCTTGAGCTCTCGCGCAGGCTC.........AAGCACC 112

 31 rgGlyProAspAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGly  47
    ::|||       ||| :::  :||  :::::|||||
113 GCGGCCCGACTGGAGCGGCATCACCAGGTCGGCTGCTACCTCTCC 162

 48 PheAsnArgLeuSerIleIleAspIleAlaHisSerHisGlnProLeuAr  64
    :::||||||:::|||||||| |||   ||||||||
163 CACCAGGCGCTCGCCATCATCGACCCTGCCTCTTGGCACCAGCCGCTC..  210

 64 gTrpGlyProAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyG  81
    ::||||         |||||||| |||||||
211 .........TACAACGAGGACAAGTCCATCGTCGTCACAGTGAATGGAG  250

 81 luIleTyrAsnTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyr  97
    ||||||||:: ::|||||| :::|||||  :::||||
251 AGATCTACAACCATGAACAGCTCCGGGCGCCAGCCTCC.....TCCCAC  294

 98 ThrPheAsnThrSerGlyAspGlyGluProIleValValGlyPheHisHi  114
    ||||||:::::::|||  ||::  |||||    :::|||||
295 ACGTTCAGGACAGGCAGCGACTGCCAGGTCATCGCACCACCTGTACGAGGA  344

114 sTrpGlyGluSerValValGluHisLeuArgGlyMetPheGlyIleAlaI  131
    |||||||  :::::::   |||::|||:  |||||||||
345 GCATGGGGAGAACTACATCGACATGCTGGATGGTGTCTTCTCCCTTCGTCT  394

131 leTrpAspThrLysGluLysSerLeuPheLeuAlaArgAspAspGlnPheGly  147
    :: ||||||:::||:||||  |||||||| ::||||
395 TGCTCGATCACGCGACACAGCTTCATTGCTGCCCGTGATGCCATTGGC  444

148 IleLysProLeuPheTyrAlaThrThrGluHisGlyThrValPhe...Se  163
    ::: |||||||:: :::|| ||::::|:: :||||:::|| |
445 GTCACACCCCTGTATTGGCTGGGGAATTGATGGGTCGGTGTGGATATC  494

163 rSerGluLysLysThrIleLeuGluMetAlaGluGluMetAsnLeu  178
```

```
     ||||||     |||    ...   :::  ::  :::||||::::::::::::
495  ATCAGAGATGAAGGGCCTGAATGATGATGATTGTGAGCACTTTGAGATC 540


seq_name: gb_est1:AV916129

seq_documentation_block:
LOCUS       AV916129             609 bp    mRNA     linear   EST 18-JAN-2002
DEFINITION  AV916129 K. Sato unpublished cDNA library, cv. Haruna Nijo
            germination shoots Hordeum vulgare subsp. vulgare cDNA clone
            bags1k02 5', mRNA sequence.
ACCESSION   AV916129
VERSION     AV916129.1  GI:18211906
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare.
ORGANISM    Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1 (bases 1 to 609)
AUTHORS     Sato,K., Saisho,D. and Takeda,K.
TITLE       Barley EST sequencing project in NIG and Okayama Univ
JOURNAL     Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..609
                     /organism="Hordeum vulgare subsp. vulgare"
                     /cultivar="Haruna Nijo"
                     /db_xref="taxon:112509"
                     /clone="bags1k02"
                     /clone_lib="K. Sato unpublished cDNA library, cv. Haruna
                     Nijo germination shoots"
                     /tissue_type="shoots"
                     /dev_stage="germination"
BASE COUNT      132 a    174 c    171 g    132 t
ORIGIN

alignment_scores:
        Quality:        243.50            Length:       182
        Ratio:          1.964             Gaps:         5
Percent Similarity:     68.132      Percent Identity:   35.165

alignment_block:
US-09-786-474-2 x AV916129

gn seg 1/1  to: AV916129   from: 1  to: 609

 1   MetCysGlyLeuGlyIleLeuThrAlaAsnGlyAsnAlaGlu...... 15
     ||||||||||||:|:|||::||||    :::   :::::::
39   ATGTGTGGCCATACTGGCAGTGCTGGGCTCGCCGATGACACCCAGGGAA 88

16   ....AlaPheValProAlaGluArgAlaLeuProCysMetArgHisA 31
     ::::::||||||||||| ||| ||:::::
89   GAGAGTGCGCGTCCTGCGCTCCGCCCAGGCTC......AAGCACC 129

31   rgGlyProAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGly 47
     ||||||||||   :::   ::: ||||||||||||||| ::::::
130  GCGGGCCCCGACTGGAGTGGCATGGACGGTGGTGACTGCTACCTCTCC 179

48   PheAsnArgLeuSerIleIleAspIleAlaHisSerHisGlnProLeuAr 64
     :::|||||||||:||||||| ::: ::  |||||||||
180  CACCAGCGCCTCGCCATCATCGACCCTGCCTCAGGCGACCAGCCACTT.. 227

64   gTrpGlyProAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyG 81
     :::|||     :: ||||||||||||| ||| |||||||
228  ........TACAACGAGGACAAGTCCATCGTCGCTCACTGTGAATGGAG 267
```

```
81   luIleTyrAsnTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyr 97
     ||||||||||||:::  ::::|||||||        :::||||
268  AGATCTACAACCATGAACAGCTCCGGGGCCCAGCTCTCC......TCCCAC 311

98   ThrPheAsnThrSerGlyAspGlyAspIleValValGlyPheHisHi 114
     |||||||::|::|||||   |||  |||:::::  :::|||
312  ACATTCAGGACAGGCAGCGACTGCGAGGTCATCGCAACCTGTACGAGGA 361

114  sTrpGluSerValValGluHisLeuArgGlyMetPheGlyIleAlaI 131
     :|||:::  :::::::::  ||| ||||||||||::::::::::
362  GCATGGGGAGAATTTCATCGACATGCTGGATGGTGTCTTCTCCTTCGTCT 411

131  leTrpAspThrLysGluLysSerLeuPheLeuAlaArgAspGlnPheGly 147
     ::: ||||||:::::::::::  |||||||| ::||||
412  TGCTCGACACCGCACACGAGCTTCATTGCTGCCCGTGATGCCATTGGT 461

148  IleLysProLeuPheTyrAlaThrThrGluHisGlyThrValPhe...Se 163
     |||||:::  |||||  :::::|||  ||:::||::||
462  GTCACACCCCTGTATGTTGGCTGGGAATTGATGGGTCCGTGTGGATATC 511

163  rSerGluLysThrIleLeuGluMetAlaGluGluMetAsnLeu 178
     ||||||  || :::|||::: :::||||||||||||||||
512  ATCAGAGATGAAGGGCCTGAATGATGATGATTGTGAGCACTTTGAGATC 557
```

OM of: US-09-786-474-2 to: Issued_Patents_NA:*   out_format : pfs

Date: Sep 20, 2002   9:32 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL-frame+.p2n.model -DEV-xlh
-Q-/cgn2_1/USPTO_spool/US0978674/runat_18092002_132912_5495/app_query.fasta_1.706
-DB-Issued_Patents_NA -QFMT-fastap -SUFFIX-p2n.rni
-GAPOP-12.000 -GAPEXT-4.000 -MINMATCH-0.100 -LOOPCL-0.000
-LOOPEXT-0.000 -QGAPOP-4.500 -QGAPEXT-0.050 -XGAPEXT-0.000
-XGAPEXT-0.000 -FGAPOP-6.000 -FGAPEXT-7.000 -YGAPOP-10.000
-YGAPEXT-0.500 -DELOP-6.000 -DELEXT-7.000 -START-1
-MATRIX-blosum62 -TRANS-human40.cdi -LIST-45 -DOCALIGN-200
-THR_SCORE-pct -THR_MAX-100 -THR_MIN-0 -ALIGN-15 -MODE-LOCAL
-OUTFMT-pfs -NORM-ext -HEAPSIZE-500 -MINLEN-0 -MAXLEN-2000000000
-SER-US09786474_@CGN1_1_54 -NCPU-6 -ICPU-3 -LONGLOG
-EV_TIMEOUT-120 -WARN_TIMEOUT-30 -NO_XLPXY -WAIT -THREADS-1

Search information block:
Query: US-09-786-474-2
Query length: 640
Database: Issued_Patents_NA:*
Database sequences: 383533
Database length: 122816752
Search time (sec): 54.020000

score_list:
Sequence	Strd	Orig	ZScore	EScore	Len		Documentation
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-125-468-1	-	633.00	1202.64	2.7e-59	30001	!	Documentation
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-474-933-1	-	633.00	1202.64	2.7e-59	30001	!	
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-090-793-12	-	463.50	855.16	5.3e-40	40138	!	
/cgn2_6/ptodata/1/ina/backfiles1.seq:5256558-3	+	425.00	816.83	8.3e-38	2002	!	Pate
/cgn2_6/ptodata/1/ina/backfiles1.seq:5256558-1	+	421.00	809.51	2.1e-37	1886	!	Pate
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-737-825-1	+	184.00	311.18	1.2e-09	8341	!	
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-385-028-16	+	165.00	294.45	1.0e-08	1542	!	
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-385-028-13	+	165.00	268.52	2.9e-07	11604	!	
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-385-028-1	+	165.00	265.15	4.4e-07	15079	!	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-599-171A-22	+	154.50	271.38	2.0e-07	1779	!	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-646-590B-22	+	154.50	271.38	2.0e-07	1779	!	
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-226-22	+	154.50	271.38	2.0e-07	1779	!	
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-412-184-22	+	154.50	271.38	2.0e-07	1779	!	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-09-328-111-544	-	152.00	280.95	5.8e-08	570	!	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-09-328-111-395	+	132.00	239.25	1.2e-05	629	!	
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-428-517-1	+	131.50	181.77	0.0195	50937	!	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-925-5	+	122.00	202.16	0.0014	2337	!	
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-902-632-3	+	118.50	195.15	0.0035	2325	!	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-09-328-111-401	-	116.50	207.23	0.0007	663	!	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-804-196-1	+	116.50	177.11	0.0354	6909	!	
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-658-340-1	+	116.50	177.11	0.0354	6909	!	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-746-111-26	+	116.50	177.11	0.0354	6909	!	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-949-637-1	+	115.00	193.78	0.0042	1491	!	
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-291-488-1	+	115.00	193.78	0.0042	1491	!	
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-328-111-668	+	114.50	203.26	0.0012	659	!	
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-328-111-456	+	112.00	200.32	0.0018	559	!	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-911-445-4	+	111.50	188.72	0.0080	1275	!	
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-182-983-4	+	111.50	188.72	0.0080	1275	!	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-09-911-445-19	+	111.50	186.99	0.0100	1458	!	
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-182-983-19	+	111.50	186.99	0.0100	1458	!	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-911-445-17	+	111.50	184.31	0.0141	1796	!	
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-182-983-17	+	111.50	184.31	0.0141	1796	!	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-911-445-5	+	111.50	183.97	0.0147	1845	!	
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-182-983-5	+	111.50	183.97	0.0147	1845	!	
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-911-445-6	+	111.50	182.64	0.0174	2046	!	
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-182-983-6	+	111.50	182.64	0.0174	2046	!	
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-375-134-13	+	111.00	177.90	0.0320	2734	!	2734
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-15263-13	+	111.00	177.90	0.0320			
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-057-570-1	+	111.00	166.32	0.1415	6736	!	
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-057-570-5	+	111.00	166.32	0.1415	6736	!	

/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-09-057-570-3 +   111.00   165.82   0.1509   7004
/cgn2_6/ptodata/1/ina/6B_COMB.seq:US-08-952-127-11 +   110.00   159.72   0.3299   9620
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-844-056-1 +   104.50   169.14   0.0985   1944
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-743-637B-6 +   104.50   164.91   0.1695   2703
/cgn2_6/ptodata/1/ina/6A_COMB.seq:US-08-526-840B-6 +   104.50   164.91   0.1695   2703

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-125-468-1

seq_documentation_block:
; Sequence 1, Application US/08125468
; Patent No. 5589385
; GENERAL INFORMATION:
.; APPLICANT: Ryan, Michael J.
.; APPLICANT: Lotvin, Jason A.
.; APPLICANT: Strathy, Nancy
.; APPLICANT: Fantini, Susan E.
.; TITLE OF INVENTION: Cloning of the biosynthetic pathway for
.; TITLE OF INVENTION: chlortetracycline and tetracycline Formation and cosmids
.; TITLE OF INVENTION: useful therein
.; NUMBER OF SEQUENCES: 1
.; CORRESPONDENCE ADDRESS:
.; ADDRESSEE: American Cyanamid Company
.; STREET: One Cyanamid Plaza
.; CITY: Wayne
.; STATE: New Jersey
.; COUNTRY: USA
.; ZIP: 07470
.; COMPUTER READABLE FORM:
.; MEDIUM TYPE: Floppy disk
.; COMPUTER: IBM PC compatible
.; OPERATING SYSTEM: PC-DOS/MS-DOS
.; SOFTWARE: PatentIn Release #1.0, Version #1.25
.; CURRENT APPLICATION DATA:
.; APPLICATION NUMBER: US/08/125,468
.; FILING DATE: 22-SEP-1993
.; CLASSIFICATION: 435
.; ATTORNEY/AGENT INFORMATION:
.; NAME: Tsevdos, Estelle J
.; REGISTRATION NUMBER: 31,145
.; REFERENCE/DOCKET NUMBER: 31,255-02
.; TELECOMMUNICATION INFORMATION:
.; TELEPHONE: (201)831-3241
.; TELEFAX: (201)831-3305
.; INFORMATION FOR SEQ ID NO: 1:
.; SEQUENCE CHARACTERISTICS:
.; LENGTH: 30001 base pairs
.; TYPE: nucleic acid
.; STRANDEDNESS: single
.; TOPOLOGY: linear
.; MOLECULE TYPE: DNA (genomic)
US-08-125-468-1

alignment_scores:
Quality: 633.00   Length: 622
Ratio: 1.829   Gaps: 20
Percent Similarity: 55.627   Percent Identity: 29.582

alignment_block:
US-09-786-474-2 x US-08-125-468-1/rev
Align seq 1/1 to reverse of: US-08-125-468-1 from: 1 to: 30001

       1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPh 17
         |||||||||||:::||:::::||:::| ||:|| ||||||
   22914 ATGTGCGGAATCTGGGGGTGGTCGACTACGACGCCCCGGCCGAACACCG 22865

      17 eValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA 34
         ||||:| |||::::||:| |||::|:|||||||||
   22864 CCGGGACCTGCTCAGGCGATGACCGACCACCATGGCCTGCCGGGGCCCG 22815

      34 spAspAlaGlyThrTrpHisAspAlaAspAlaAlaGlyPheAsnArg 50
         ||| ||| :::::||| .:|||| ..::|||

```
323 ValLysIleValSerProGluGluTyrAlaAsnAlaIleProLysIleMe 339
        |||    |||||
21948 GTGCTGCTGGACAGCCCGGC............................ 21928

339 tTrpTyrLeuAspAspProValAlaAspProSerLeuVal.......... 352
        |||:|||||||
21927 ......CTGTCGGACCCCGGCGGTGCGTTCCGGCGTGCTGCCGCCACCG 21885

353 .................................ProLeuTyrPheVal 357
                                          |||||||:::
21884 ACCTGCGCCCCGCTACTGGGCGACCATGTGGCCCTTCGCTCACCTGTTC 21835

358 AlaAlaGluAlaArgLysHisValLysValLeuSerGlyGluGlyAl 374
        ||:|||||||||||||||||||||||||||||||||||||||:||
21834 TTCCCGCCCGTGCGGGAGAACGTCACGGTGGCCGGTCCCGGAGGCCGC 21785

374 aAspGluLeuPheGlyGlyTyrThrIleTyrLysGluProLeuSerLeu. 390
        ||||||||||||||:::    |||||||||||||||||||:||:|||
21784 GGACGAACTTCTTCGGCGGCTACCGTGGTTCCGAACCCGTTGGCCCTGC 21735

391 ....AlaProPheGluLysProSerProLeuArgLysGlyLeuGly 405
                  |||||        |||    ||
21734 GGGCGCGGCCACCTTC................................ 21721

406 LysLeuSerLysValLeuProAspGlyMetLysGlyLysSerLeuLeuGl 422
                                          |||
21720 ............................................CCGTGGCTCAC 21710

422 uArgGlySerMetThrMetGluGluTyrTyrGlyAsnAlaArgSerP 439
        ||:|||                |||:|||||||::||:|||
21709 CCCGGTTCG.....GGCGGCTACTCGGCGGCAGCTCGCTGT 21672

439 heAsnPheGluGlnMetGlnArgValIleProTrpAlaLysArgGluTrp 455
        ||:        :::||||||||
21671 TCCAGGCGGGTCTCTGGACAAGCTCGACCTGGACGGCGCCGGCGCC 21622

456 AspHisArgGluValThrAlaProIleTyrAlaGlnSerArgArgAsnPheAs 472
        :::|||||||||||||||||||:|||||||||||||||
21621 CGGTACGGCCGAGGCCGCTCGCCGAAGTGCCGGTGCTGCCGGTAGTCGGC 21572

472 pProValAlaArgMetGln.........HisLeuAspPheThrTrpM 486
        |||                          |||||||||:|||
21571 GCCCGAGAGCGGGATGCGGGAGATCGGTCACCTGAACCTGACCCGGTTCG 21522

486 etArgGlyAspIleLeuValLysValAlaAspLysIleAsnMetAlaAsnSer 502
        :::|||    :::    |||    :::||||||||||||
21521 TGCAG......ACCCTGCTCGACAACGCCAAGGACCGGATGAGCAGTCCGGTCGGC 21475

503 LeuGluLeuArgValProPheLeuAspLysLysGluValValAlaGl 519
        ||||||||||||:|||||||||||    :::|||
21474 CTCGAGGTGCGGGTGCCGTTCTCGGACAAGAAGGAAGTGGTGGCCGGG 21425

519 uThrIleProTyrAspProLeuLysIleAlaAsnGlyThrThrLysTyrAlaL 536
        :::|:|||||||||||:::|||||    ::|||
21424 CAACGGCCATCCCGTACGACCCGCTGAAGATCGCCAACGGGACGACCAAGTACGCCCTGC 21375

536 euArgArgAlaLeuGluGlnIleValProHisValLeuHisArgLys 552
        |||||||||||||||||||||||||||||||||||||
21374 TGCGGCGGGCCCTGGAGCAGATCGTGCCGCACGTGCTGCACCGGAAG 21325

553 LysLeuGlyPheProVal.......................ProMetAr 561
        |||    |||:
21324 AAGACCCCGTACCCGGCGACCCCAGGATCCGGCTACGAGGGCGCCCTGCG 21275

561 gHisTrpLeuAlaGly 566
        |||||||||||
21274 GGCCGAACTGGCCCGG 21259

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-474-933-1

seq_documentation_block:
; Sequence 1, Application US/08474933
```

```
22814 ACCCGGAGGGCCTCTGGTTCGGCGCCGGCGGGTCGGCTCACCGCGGC 22765
        ||||||||||||||||||||||||||||||||||||||||||
51 LeuSerIleLeuAspIleAlaHisSerHisGlnProLeuArgTrpGlyPr 67
        ||||||||||||||||||||||||||||||||||:::::::|||||
22764 CTGTCGGTGATCGACCCCGAGCACGGCCGGCAGCCGATG.......AC 22724

67 oAlaAspGluProAspArg...TyrAlaMetThrPheAsnGlyGluIleT 83
        |||            |||||||||||||||||||||||||||||
22723 CGCGAACACGAGGCCGCCGGCGGGCCGGATCACCTTCAGCGGCGAGATCT 22674

83 yrAsnTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyrThrPhe 99
        :::|||||||    ||||||||    |||||||:::::
22673 ACAACTTCCGTGAGCTGCGCGGCGAACTCACCTGCCACGGCCACCGGTTC 22624

100 AsnThrSerGlyAlaAspTyrGlyIleValValGlyPheHisHisTrpGl 116
        |||||||    |||    ||||||    :::|||
22623 AGGACGTCCTGCGATACCGAGGTCGTCCTGCGCGGCTACCTCCAGTGGGG 22574

116 yGluSerValValGluHisLeuArgGlyMetPheGluIleAlaIleTrpA 133
        :::|||||||||||||||||||||||||||||||||:::||
22573 CGCCGGCTGGTCGAGAGGCCTCAACGGGATGTTCGCCTTCGCCGTCTGGG 22524

133 spThrLysGluLysSerLeuPheLeuAlaArgAspGlnPheGlyIleLys 149
        |||:::||||||||||||||||||:::|||||:::
22523 ACCAGCACAGCGAGGAACTCCTGTCCGCGACCCGGATGGGCGTCAAG 22474

150 ProLeuPheTyrAlaThrThrGluHisGlyThrValPheSerSerGluLy 166
        ||||||    |||    :::|||    |||:::
22473 CCGCTCTACTACTGGCCGACCGTGCGCTTCGGCTCGGAGCC 22424

166 sLysThrIleLeuGluMetAlaGluAsnLeuAsnLeuGlyLeuGlyLeuA 183
        |||||||                        |||:::
22423 CAAGGCCCTCCTG..............................GCCGACCGGGACCCTGT 22395

183 spLysArgThr.............IleGluHisTyrValAspLeuGlnTyr 195
        |||:::                  :::||||||||||
22394 CCCGCCGGGTCGGCCCCGACGGGTGCGAGGTGCTGGACAATGGTGAAG 22345

196 ValProGluProAspThrLeuHisAlaGlnIleIleSerArgLeuGluSerGl 212
        |||||||||    :::    :::
22344 ACGCCCGAGGCGGCCGGTCGTTCTCCGGGCTCGGAG.............. 22309

212 yCysThrAlaThrValAlaArgProGly.................GlyL 222
        |||||||||||||||||||
22308 .........GTCCTCCGGTCAGCTCCGCTCCGGTGGGGCGCCGGCG 22272

222 ysLeuGluGlnLysArgTyrPheLysGlnInPheProValGlnLysVal 238
        |||:::|||||||    |||:|||    :::|
22271 GGGTGCGCCGCAGACGTACTGGCGGCTGGAG.................. 22240

239 ValLysGlyLysGluGlnAspLeu.........PheAspArgIleAlaGl 252
        :::||||||||||            :::|||
22239 GCCCGAGGAGCAGACCGACCGCCGCACGACCATCGGCACCGTCCGTGA 22190

252 nValLeuGluGlnLysSerValGluLysHisMetArgAlaValThrValG 269
        ::::||||||||||||||:::||||    :::||
22189 CCTGCTCGCCCGACACCGTGCTCCAGCTGCGTTCCGGGCTGG 22140

269 lySerPheLeuSerGlyIleGlyIleIleAspAspSerThrAlaIleAlaProLeuAla 285
        |||||||||:::||:||    ||||||    |||
22139 GCACCGCTGCTCTCCGGCATCGGCATCATCGGCGACTCCGTGACGGCCGCTGGCC 22090

286 LysArgHisAsnProAspLeuTyrTrpHisProGlyThrArgGlyPhe....... 299
        |||:::    :::|||||||||||||
22089 GGCCCGCAGCGGCCG...GTGCGCTCCTTCTCCGACTGGGACTTCCGGGCGGCGGC 22043

300 ......................GluArgGluGlyTyrSerGluLuV 307
                              |:::||||||    :::
22042 CGCCGACGCCCTTCGCCCTCGACGGCGGGAG.......GCGGACG 21999

307 aIleAspValAlaAlaGluSerAlaAlaAlaAlaIleGlyAlaGluHisIle... 322
        ::    |||    :::    |||||||||||||||||
21998 CGCCGTACGTGCGCGAGCTGCGGGGCACGTCGGAGCCGGACCACCTGGAG 21949
```

; Patent No. 5866410
; GENERAL INFORMATION:
;   APPLICANT: Ryan, Michael J.
;   APPLICANT: Lotvin, Jason A.
;   APPLICANT: Strathy, Nancy
;   APPLICANT: Fantini, Susan E.
;   TITLE OF INVENTION: Cloning of the biosynthetic pathway for
;   TITLE OF INVENTION: chlortetracycline and tetracycline Formation and cosmids
;   TITLE OF INVENTION: useful therein
;   NUMBER OF SEQUENCES: 1
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: American Cyanamid Company
;   STREET: One Cyanamid Plaza
;   CITY: Wayne
;   STATE: New Jersey
;   COUNTRY: USA
;   ZIP: 07470
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/474,933
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/125,468
;   FILING DATE: 22-SEP-1993
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Tsevdos, Estelle J
;   REGISTRATION NUMBER: 31,145
;   REFERENCE/DOCKET NUMBER: 31,255-02
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (201)831-3241
;   TELEFAX: (201)831-3305
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 30001 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
; US-08-474-933-1

lignment_scores:
        Quality: 633.00   Length: 622
        Ratio: 1.829   Gaps: 20
Percent Similarity: 55.627   Percent Identity: 29.582

alignment_block:
US-09-786-474-2 x US-08-474-933-1/rev ..

Align seg 1/1 to reverse of: US-08-474-933-1 from: 1 to: 30001

```
; CURRENT FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/048,650
; EARLIER FILING DATE: 1997-06-04
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;  LENGTH: 40138
;  TYPE: DNA
;  ORGANISM: Vibrio marinus
US-09-090-793-12
```

```
alignment_scores:
        Quality:     463.50                    Length:      589
        Ratio:       1.525                      Gaps:         20
Percent Similarity:  51.613     Percent Identity: 25.127
```

```
alignment_block:
US-09-786-474-2 x US-09-090-793-12

Align seg 1/1   to reverse of: US-09-090-793-12   from: 1   to: 40138
```

```
       1 MetCysGlyIleLeuGlyIleIleLeuThrAlaAsnGlyAsnAlaGluAlaPh 17
         ||||||::|| :::     ||||::::   :::   :: |||
   11291 ATGTGCTCAATATTCGGAATTCTAGACATCAAATCAGACATTAAA......  11247

      17 eValProAlaLeuUArgAlaLeu.............ProCysMetArgHisA 31
         |||  :::::|| :|||                       ||||||
   11246 ....CCTCTTCGCGACAAGCACTAGAACTATCAAAGTTACTACGTCACC  11201

      31 rgGlyProAspAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGly 47
         |||||||||| :|:    |||::|::::::  ||   :: |||
   11200 GCGGGCCCAGATTGGTCAGGTATCTTATACAAGTGACAATGCTATTTTAGTT 11151

      48 PheAsnArgLeuSerIleIleAspIleAlaHisSerHisGlnProLeuAr 64
         |||||||||  :: ||||   : ::      :|||||||||||||
   11150 CATGAACGTCTTGCTATTGTCGATGTAAATAATGGTGCACAGCCGCTA..  11103

      64 gTrpGlyProAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyG 81
         ||||             :::: |||     :::|| ||||
   11102 .........TATAATGAAGAAAAGACACACGTACTTGCTGTTAATGGTG  11063

      81 luIleTyrAsnTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyr 97
         |||||||||||||::      ||::  :||       ::|||
   11062 AAATTTATACCACAAAGATTTAAAGAAAACGCTG...AACGTAGACTTT 11016

      98 ThrPheAsnThrSerGlyAspGlyGluProIleVaIGlyPheHisHisHi 114
         ::::|||     |||            :::|||::::||||||
   11015 GAATTCCAAACAGAATCAGATTGCCGAAGTGATACTTGCGCTATATAAAGA 10966

     114 sTrpGlyGluSerValValGluHisLeuArgGlyMetPheGlyIleGluIleAlaI 131
         |||   :::::|     ||||::||||::: ||||||
   10965 AAAAGGCACACAATTTTAGATGATCTAAATGGTATTTTTGCTTTCGCTT 10916

     131 leTrpAspThrLysGluLysSerLeuPheLeuAlaLeuArgAspGlnPheGly 147
         ::::|||   ||| ||||||||||:|||||:::||||
   10915 TATATGACGAAACAGAGGACGCCTTACCTGATTGGCCGTGACCATATCGGT 10866

     148 IleLysProLeuPheTyrAlaThrThrGluHisGlyThrValPhe...Se 163
         |||  |||||::::||  :: ||||||  :|| |||||||
   10865 ATCATCCCTCTTTATACGGGCTATGATGAACACGTAACTTTTATGTTGC 10816

     163 rSerGluLysLysThrIleLeuGluMetAlaGluGluMetAsnLeuAspL 180
         ::|||||   |||:::::|||       ::|
   10815 CTCTGAAATGAAAGCACTAGTGCCAATTTGTACTCAA......  10779

     180 euGlyLeuAspLysArgThrIleGluHisTyrValAspLeuGlnTyrVal 196
         ||||||   :::::||  ||| ||||   ::||| ::||||
   10778 ....ATTGAAGAATTCCCAGCCAGTCATTACCTG 10749

     197 ProGluProAspThrLeuHisAlaGlnIleSerArgLeuGluSerGlyCy 213
         |||||||||  ||| ||| |||   ||||||::|||||
   10748 TGGAGTAAGGAT...........................  10737
```

Left column alignment data:

```
   21927 ......CTGTCGGACCCGGCGGGTGCGTTCCGCGGGTGCTGCGCCACCG 21885
     353 .................................ProLeuTyrPheVal 357
                                          |||||:::|:
   21884 ACCTGCGCGCCCCGTACTGGGGCGACATGTGGCCCTGCTCTTACCTGTTC 21835

     358 AlaAlaGluAlaArgLysHisValLysValLeuSerGlyValGluGlyVal 374
         ::||:||: ||:|||||   ||||||||::::|||
   21834 TTCCGGCGGTGCGGGAGAACGTCACGTGGCCGTGCCGGGCGAGGCCGGC 21785

     374 aAspGluLeuPheGlyTyrThrIleTyrLysGluProLeuSerLeu. 390
         ||||||||||||||   :::|||:: ::|||||
   21784 GGACGAACTCTTCGGCGCGCTACCGGTGTTCCCAACCCGTTGGCCCTGC 21735

     391 ....AlaProPheGluLysIleProSerProLeuArgLysGlyLeuGly 405
             ||| |||:::||
   21734 GGGCGGCCACCTTC.................................. 21721

     406 LysLeuSerLysLeuValProAspGlyMetLysGlyMetLysSerLeuLeuGl 422
                                                    |||
   21720 .........................................CCGTGGCTCAC 21710

     422 uArgGlySerMetThrMetGluGluArgTyrTyrGlyAsnAlaArgSerP 439
         |||||                :::
   21709 CCCGGGTTCG.........GCGCCGCTACTTCGGCGCAGCTCGCTGT 21672

     439 heAsnPheGluGlnMetGlnArgValIleProTrpAlaLysArgGluTrp 455
         :::::||||   |||
   21671 TCCAGGCGGGTTGCTGCCAAGCTGACCTGACAGCCGTACCGGGCGGCGCC 21622

     456 AspHisArgGluValThrAlaProIleTyrAlaGlnSerArgAsnPheAs 472
         ::: ||| |||
   21621 CGTACGCCGAGGCGCTGCCCGAAGTGCGCGGTGCTGCCGGGGTGAGTCGGT 21572

     472 pProValAlaArgMetGln..........HisLeuAspPheThrTrppM 486
         ||||||          :::      |||||
   21571 GCCCGAGAGGCCGGATGCGCGAGATCGGTCACCTGAACCTGACCCGCTTCG 21522

     486 etArgGlyAspIleLeuValLysAlaAspLysIleLeuAsnMetAlaAsnSer 502
         :::::|| :::      |||::::|||
   21521 TGCAG...ACCCTGCTCGCACCGCAAGGACCGGATGAGCATGGCGGTCGGC 21475

     503 LeuGluLeuArgValProPheLeuAspLysValPheLysValAlaAlaGl 519
         ||||||:::|||||||||   ::: :::
   21474 CTCGAGGTGCGGGTGCCGTTCTGCGACCACCGCTCGTGGTACGTCTT 21425

     519 uThrIleProTyrAspLeuLysIleAlaAsnGlyThrThrLysTyrAlaL 536
         :::::||  ||    |||       ||
   21424 CAACGTGCCCTGGGAGATGAAGAGTTTCGACGGCCGGGAGAAGAGCCTGC 21375

     536 euArgArgAlaLeuGluGlnIleLeuProProHisValLeuHisArgLys 552
         |||::||                    |||   :::|||
   21374 TGCGGGGCGGCGGTGGCGCACCTGCGCGCGCTGCGGGTGCTCGGGGTG 21325

     553 LysLeuGlyPheProVal.............ProMetAr 561
         |||   :::|||:::                   ::|||
   21324 AAGACCCCGTACCCGGCGCCAGGATCGGGCTACGAGGGCGCCCCTGCG 21275

     561 gHisTrpLeuAlaGly 566
         ||||||||
   21274 GGCCGAACTGGCCGGG 21259
```

```
seq_name: /cgn2_6/ptodata/1/1na/6A_COMB.seq:US-09-090-793-12
```

```
486 tArgGlyAspIleLeuValLysAlaAspLysIleAsnMetAlaAsnSerL 503
    |||:::|||                            |||    :::
10131 ACGT.....................GCAAACAAATCTATGGCAGCTTGGGGCA 10100

503 euGluLeuArgValProPheLeuAspLysGluValPheLysValAlaGlu 519
    :|||    |||||||||||||||||||||
10099 TCGAAGCACGTGTTCCTTTCTTAGATAAAGAATTTGTGATGCGTCAATG 10050

520 ThrIleProTyrAspLeuLysIleAlaAsnGly......ThrThrLysTy 534
    :::|||||||:::|||  :::|||                 |||
10049 CGTCTAAACCCTGAACTAAAAATGATCACGGGCGACCGCATTGAGAAAAA 10000

534 rAlaLeuArgArgAlaLeuGluGluGlnIleValProProHisValLeuHisA 551
    :::|||::::|||::|||::|||::|||||||    :::|||          |
9999 CATCATTCGAGAAGCATTTGAAGCATTTACTACCTGAAGAAATTGTATGGC 9950

551 rgLysLysLeuGlyPhe 556
    |||:::
9949 GTCAAAAAGAACAATTC 9933


seq_name: /cgn2_6/ptodata/1/ina/backfiles1.seq:5256558-3

seq_documentation_block:
;Patent No. 5256558
;APPLICANT: CORUZZI, GLORIA M.; TSAI, FONG-YING
;TITLE OF INVENTION: GENE ENCODING PLANT ASPARAGINE SYNTHETASE
;NUMBER OF SEQUENCES: 17
;CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/07/514,816
;  FILING DATE: 26-APR-1990
;PRIOR APPLICATION DATA:
;  APPLICATION NUMBER: 347,302
;  FILING DATE: 03-MAY-1989
;SEQ ID NO:3:
;  LENGTH: 2002
5256558-3

alignment_scores:
         Quality:   425.00           Length:    592
         Ratio:     1.384            Gaps:       24
Percent Similarity: 51.858       Percent Identity: 25.845

alignment_block:
US-09-786-474-2 x 5256558-3
Align seg 1/1   to: 5256558-3     from: 1  to: 2002

1 MetCysGlyLeuLeuGlyIleLeuThrArgAlaAsnGlyAsnAlaGluAla.. 16
  |||||||||:::|||::::|||::|||     ::::|||
102 ATGTGTGGTCTATACTTGCTCGTTCTTGGTTGTTCGATCCTTCTCGAGCCAA 151

17 .......PheValProAlaLeuGluArgAlaLeuProCysMetArgHisA 31
   ::|||||||||:::|||
152 GAGAGTTCGTGTGTTGGAACTTTCACGCAGATTG.........AAGCACC 192

31 rgGlyProAspAspAlaGlyThrTrpHisAspAlaAspAlaAlaPheGly 47
   |||||||    :::|||
193 GAGGCCCTGAATGGAGTGGGCTCCACCAACATGGTGATTGTTATTTGGCA 242

48 PheAsnArgLeuSerIleIleAspIleAlaHisSerHisGlnProLeuAr 64
   |||||||
243 CAACAACGGTTAGCCATAGTTGATCCTGCTTCGTGGTGATCAACCTCTC.. 290

64 gTrpGlyProAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyG 81
   :|||||||
291 .......TTCAATGAGACAACATCCGTCAATTGTCACGGTAAACGGAG 330

81 luIleTyrAsnTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyr 97
   |||||||:::|||
331 AGATTTACAATCATGAAGATCCTCAGGAAAACAGTTGTCTAAT.....CAC 374

98 ThrPheAsnThrSerGlyAspGlyAspGluProIleValValGlyPheHisHi 114
```

```
213 sThrAlaThrValArgProGlyGlyLysLeuGluGlnLysArgTyrPheL 230
    :::|||::|||  |||                          .......TATTATC 10715
10736 ....GGTGTAGTCACGCCT.........................

230 ysProGlnPheProValGlnLysValLysGlyLysGluGlnAspLeu 246
    :: :|||||||||||  :|||::...............CGTGATTGGAAAGAT 10698
10714 AA........

247 PheAspArgIleAla..................GlnVa 253
    |||||:||||:::|||                    |||
10697 TTTGATAACGTTGCCACAAAATGGTGATAAAAGCGGTTGTTAAGCAAGG 10648

253 lLeuGluAspSerValGluLysHisMetArgAlaAspValThrValGlyS 270
    :|||||||||  ::|||::::|||||||::......  :|||||  |||
10647 TTTGGAAGATGCAGTTAAACGTCAATTAATGTGATGTACCTTACGGCG 10598

270 erPheLeuSerGlyGlyIleIleAspSerThrAlaIleAlaProLeuAlaLys 286
    :::||||||||||:::|||::......  :::|||
10597 TATTATTATCAGGCGGGTTAGATTCATCGTGTTATTTCAGCAATCACCCAA 10548

287 ArgHisAsn................... 289
    :::::::
10547 CAGTATGCAAAACGTCGTATTGAAGACGTGGTAAAACCGAAGCATGGTG 10498

290 .ProAspLeuThrPheThrThrGlyPheGluArgGluGlyTyrSerG 306
    ::|||||||::::|||  :::|||............AATGGTTCTCCAG 10454
10497 GCCACAACTACACTCTTTCTGTAGGCCTA.....

306 luValAspValAlaLeuAlaAlaGluSerAlaAlaAlaIleGlyAlaGluHisIle 322
    ::|||::|||  |||                      :::|||
10453 ATTTAGCCCGTGCACAAAAAGTGGCAGATCACTTAGGTACTATCCATCAT 10404

323 ValLysIleValSerProGluGluTyrAlaAsnAlaIleProLysIleMe 339
    :::|||::......:::|||::|||
10403 TCAATTGAATTACTGTGCGAAGAAGGTATCGACGCTTTACGCGACGTGAT 10354

339 tTrpTyrLeuAsp.......AspProValAlaAspProSerLeuValProL 354
    :::|||.....||:|||
10353 TTACCACATTGAAACCTACGATGTAACAACAATTCGTGCATCAACCCCTA 10304

354 euTyrPheValAlaAlaGluAlaArg..LysHisValLysValValLeu 369
    ::|||:||:  :::.... :::|||||||:::|||||
10303 TGTACCTAATGGCACGTAAAATTAAAGCCATGGGATTAAGATGTGCTT 10254

370 SerGlyGlyAlaAspGluLeuPheGlyGlyTyrThrIleTyrTyrLysGl 386
    |||||||||||::|||||||||||||||  :::|||
10253 TCTGGTGAAGGTGCTGATGAACTGTTCGGTGGTTAC...TTGTAC..... 10212

386 uProLeuSerLeuAlaProPheGluLysIleProSerProLeuArgLysG 403
    :::|||::.... :::|||::|||
10211 ..........TTCCACAAAGCGCCA.......... 10197

403 lyLeuGlyLysLeuSerLysValLeuProAspGlyMetLysGlyLysSer 419
    ||||||||||:::|||
10197 GAAAGAATTCCATGAAGAGACGTACGTAAAGTA.......... 10158

420 LeuLeuGluArgGlySerMetThrMetGluArgTyrTyrGlyAsnAl 436
    |||||
10196 ..........AATGC 10192

436 aArgSerPheAsnPheGluGlnMetGlnArgValIleProTrpAlaLysA 453
    :|||||||::|||
10191 ..........TTCCACAAAGCGCCA.......... 10158

453 rgGluTrpAspHisArgGluValThrAlaProIleTyrAlaGlnSerArg 469
    :::|||
10158 ..........

470 AsnPheAspProValAlaArgMetGlnHisLeuAspPheThrTrpPheMe 486
    |||  ||:|||
10157 ..........AATAAATTACATATGTTTGATTGCTT 10132
```

```
       ||||||:::|||                    :::|||
375  ACGTTTAGGACCGGAAGTGATTGGTGATGTTATTGCCATTTGTACGAGGA  424
114  sTrpGlyGluSerValValGluHisLeuArgGlyMetPheGlyIleAlaI  131
     ::||||||:::    |||::|    |||   |||::||||||||::::|
425  ATATGGAGAAGACTTTGTGGATATGTTGGATATATTTTCGTTTGTTC     474
131  leTrpAspThrLysGluLysSerLeuPheLeuAlaArgAspGlnPheGly  147
     |||||||:::||||||||::||||:||     |||||||    ::|||
475  CATTGGATACTCCTGACAACAGTTATATTGTGGCTAGAGATGCGATTGGT  524
148  IleLeuLysProLeuPheTyrAlaThrThrGluHisGlyThrValPhe...Se  163
     ::||||    :::                    |||::::||||    ||
525  GTAACTTCTCTATACAATGGTGGGATTAGGATGGTTCGGTTTGGATTTC   574
163  rSerGluLysLysThrIleLeuGluMetAlaGluGluMetAsnLeuAspL  180
     ||||||  |||
575  GTCGGAAATGAAA...                                     587
180  euGlyLeuAspLysAspThrIleGluHisTyrValAspLeuGlnTyrVal  196
     ||||||||          :::
588  .GGTTTGAACCAT...GATTCTGAACATTTC...       GAGTGTTTT   623
197  ProGluProAspThrLeuHisAlaGlnIleLeuSerArgLeuGluSerGlyCy  213
     |||  ||||||::|:      ::|||
624  CCA...CCTGGTCATTTGTAT........TCGAGCAAAGATAGTGGCTT    661
213  sThrAlaThrValArgProGlyGlyLysLeuGluGlnInLysArgTyrPheL  230
                     :::|||::|||::|
662  T.....             AGAAGATGGTATA                     675
230  ysProGlnPheProValGlnGlnValValLysGlyValValLysGlyGlnAspLeu  246
     ::|||    |||::     :::|||    |||::|
676  ATCCTTCTTGGTACTCTGAGCTATTCCGTCCGGCTCCTTATCGATCCTCTT 725
247  PheAspArgIleAlaGlyGlnLeuLeuLeuGluAspSerValGluLysHisMetAr  263
     ::     :::::|||:::|||    |||::|||
726  ......GCTTTGAGGCACGCCTTCGAAGGCGGTGTAAAAAGGTTGAT      769
263  gAlaAspValThrValGlySerPheLeuSerGlyGlyIleAspSerSerThrA  280
     ::|||||    |||   ||||||||||||||:::|||||||||:::
770  GACAGATGTACCTTCGGTGTTCTACTATCCGGAGGTTTGGACTCGTCAT    819
280  laIleAlaProLeuAlaLysArgHis.........................  288
     ::|||||    |||::|:
820  TGGTTGCATCCATCACTTCTCCGCTACCTAGCAACCACGAAAGCGGCTGAA  869
289  ......AsnProAspLeuLeuThrPheThrPheGlyPheGluArgGluGl  303
             ::|||   ::|||:::|||   :::|||:::        |||||
870  CAATGGGATCAAAACTTACCATTCATTCTGCGTTGGACTC......GAGGG  913
303  yTyrSerGluValAspValAlaGluSerAlaLeuAlaIleGlyAlaG     320
     |   ::|||    :::||:||   |||   |||::|||::::|
914  CTCACCTGGTCTTAAGGCTGGAAAGAAGTTGCAGATTATCTCGGAACCG   963
320  luHisIleValLysIleValSerProGluGluTyrAlaAsnAlaIlePro  336
     ::     ::|||   :::|||   ||||||
964  TTCATCATGAGTTTACCTTTACTGTTCAGGATGGTATAGATGCAATTGAG  1013
337  LysIleMetTrpTyrLeuAsp......AspProValAlaAlaAspProSerLe 351
             |||   :::|||        |||   :::
1014 GATGTTATATACCATGTTGAAACATATGATGTTACTTCAATTAGAGCAAG  1063
351  uValProLeuTyrPheValAlaAlaGluAlaArgLys..HisValLysv  367
     :::||:||   :::::|||  :::  |||::        ||||||
1064 CACGCCTATGTTTCTCATGTCGAGGAGAGATTAAATCACTTGGTGTCAAAT 1113
367  alValLeuSerGlyGlyLysGluGlyLyaAspAspLeuPheGlyIleTyrThrIle 383
     :::|||||||||||||||||||||||||||||:::        ::|||
1114 GGGTGATCTCCGGTGAAGGATCCGATGAGATCTTTGCGGATAT...CTG  1160
384  TyrLysGluProLeuSerLeuAlaProGluGluLysIleProSerProLe  400
     |||                       |||
```

```
1161 TAC.......................TTTCACAAGGCACCG........   1178
400  uArgLysGlyLeuGlyLysLysLeuSerLysValLeuProAspGlyMetLysG 417
1178 ................................................     1178
417  lyLysSerLeuLeuArgGlySerMetGluArgTyrTyr               433
1178 ................................................     1178
434  GlyAsnAlaArgSerPheAsnPheGluMetGlnArgVallleProTr      450
     |||        :::|||:::     |||
1179 ...AACAAGGAAGAGTTCACGAAGAAACTTGCCGCAAGATC......AA   1219
450  pAlaLysArgGluTrpAspHisArgGluValThrAlaProIleTyrAlaG  467
     |||                ::::::|||   :::
1220 AGCACTGCACCAATATGATTGCCAG...                         1244
467  lnSerArgAsnPheAspProValAlaArgMetGlnHisLeuAspLeuPhe  483
1244 ................................................     1244
484  ThrTrpMetArgGlyAspIleLeuValLysAlaAspLysIleAsnMetAl  500
     :::::|||                    ::::::|||::|
1245 ......................AGAGCTAATAAATCGACTTATGC        1267
500  aAsnSerLeuGluLeuArgValProPheLeuAspLysGlyGluValPheLysV 517
     ::|||||   ||||||||||||||||||||::::::|
1268 TTGGGGTTTAGAAGCTAGAGTTCCGTTTCTGGACAAGGGCGTTTATCAATG 1317
517  alAlaGluThrIleProTyrAspLeuLysLysIle............AlaAsn 529
     |||||   :::|||       |||::
1318 TTGCGATGAATATTGATCCTGAGAATAAAATGATAAAACGAGATGAAGGA  1367
530  GlyThrThrLysIleProTyrAlaLeuArgArgAlaLeuGluGln.......... 542
     :::|||::|||||||||||||||||::|||
1368 CGAATTGAGAAGTATATTTTTGAGGAAGGCCATTTGATGACCGAAGAGAATCC 1417
543  .IleValProProHisValLeuHisArgLysLysLeuGlyPheProValP  559
     :::|||    :::|||   ||||||::::::|||
1418 TTATCTGCCAAAGCACACATTTTGTATAGGCAGAAGAACAATTCAGTGATG 1467
559  roMetArgHis...TrpLeuAlaGly                           566
     :::|||        :::|||     |||
1468 GAGTTGGTTATAGCTGGATTGATGGT                           1493
```

seq_name: /cgn2_6/ptodata/l/lna/backfiles1.seq:5256558-1

seq_documentation_block:
; patent No. 5256558
; APPLICANT: CORUZZI, GLORIA M.;TSAI, FONG-YING
; TITLE OF INVENTION: GENE ENCODING PLANT ASPARAGINE SYNTHETASE
; NUMBER OF SEQUENCES: 17
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/514,816
;   FILING DATE: 26-APR-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 347,302
;   FILING DATE: 03-MAY-1989
;SEQ ID NO:1:
;   LENGTH: 1886
5256558-1

alignment_scores:
          Quality: 421.00   Length: 589
          Ratio: 1.427   Gaps: 22
Percent Similarity: 50.085   Percent Identity: 25.637

alignment_block:
US-09-786-474-2 x 5256558-1

Align seg 1/1 to: 5256558-1   from: 1   to: 1886

```
285 ..........................AlaLysArgHisAsnProAspLeuLeuThrPh 295
                                   ||||||  :::  ::  ::|||
841 CTTGCTGGTACAAAAGCTGCTAAGCACAGTGGGAGCAAAATTGCCCTCTTT 890

295 eThrThrGlyPheGluArgGluGlyTyrSerGluValAspValAlaAlaG 312
    :::|||:|||::  ||:||||     :::::::| |  :::::::|
891 CTGTGTAGGCCTT......AAGGCGGCACCTGACCTAAAGGCTGGAAAGG 934

312 luSerAlaAlaAlaIleGlyAlaGluHisIleValLysIleValSerPro 328
    || |||||||||  ::::  || ||:|||   |||::::||||||||||
935 AGGTAGCAGATTTCTTAGGAACTGTCCATCATGAATTTGAGTTTACTATC 984

329 GluGluTyrAlaAsnAlaIleProLysIleMetTyrTyrLeuAsp...... 343
    :::::: ::||||||||||||   :::::::||||||||| ::
985 CAGGACGGTATAGATGCAATTGAAGATGTCATCTATCAACAGAAACATA 1034

344 .AspProValAlaAspProSerLeuValProLeuTyrPheValAlaAlaG 360
      |||                ||||||||||||||||||||||||||
1035 TGATGTTACTACGATAAGGGCTCCAACACCTATGTTTCTGATGTCTCCTA 1084

360 luAlaArgLysHis...ValLysValValLeuSerGlyGluGlyAlaAsp 375
    :: :|||||||||    ||||||  ||||||||||||||||||||:|||
1085 AGATCAAATCATCCGAGTCAAATGGGTGATTTCTGGAGAGGATCTGAT 1134

376 GluLeuPheGlyGlyTyrThrIleTyrTyrLysGluProLeuSerLeuAlaPr 392
    |||::|||||||||||||   ::::||
1135 GAGATCTTTGGAGGGTAT...TTGTAT.................... 1158

392 oPheGluLysIleProSerProLeuArgLysGlyLeuGlyLysLeuSerL 409
    |||||||||||||||||||||||||:::||
1159 .TTCCATAAGGCGCCAAAC............................ 1176

409 ysValLeuProAspGlyMetLysGlyLysSerLeuLeuArgGlySer 425
    ::..................................................... 1176

426 MetThrMetGluGluArgTyrTyrGlyAsnAlaArgSerPheAsnPheGl 442
    ::................................................... 1176

442 uGlnMetGlnArgValIleProTrpAlaLysArgGluTrpAspHisArgG 459
                                      |||||||  ::|||:||
1177 ...............................AGGGAAGAGTTTCACCAAG 1195

459 luValThrAlaProIleTyrAlaGlnSerArgAsnPheAspProValAla 475
    ||:::        ||||||::|||  || :::|||
1196 AAACATGCCGCAAGATCAAAGCTCTTCATAGA...TATGAT......... 1233

476 ArgMetGlnHisLeuAspLeuPheThrTrpMetArgGlyAspIleLeuVa 492
    .................................................TGTTT 1238

492 lLysAlaAspLysIleAsnMetAlaAsnSerLeuGluLeuArgValProP 509
    :::||||:||  ::::    :: ||||::  ||||::||::||||
1239 GAGAGCCCAATAAATCAACATATCATGGGGTCTAGAAGCTAGAGTACCAT 1288

509 heLeuAspLysGluValPheLysValAlaGluThrIleProTyrAspLeu 525
    |||||||||  ::::||||||||||||||||||  ||  :::::::
1289 TTTTGGACAAGGACTTTATCAAGGTTGCAATGGACATTGATCCTGAGTTT 1338

526 LysIle.............AlaAsnGlyThrThrLysTyrAlaLeuArgAr 538
    |||:::             :::  ||||:||  |||||
1339 AAAATGATAAAACATGATGAAGGAAGAATTGAGAAATTGGATTCTAAGAAA 1388

538 gAlaLeuGluGln.............IleValProProHisValLeuHisA 551
    ::|||            |||||  ::::|||:::|||::|:
1389 GGCCTTTGATGATGAAGAGAATCATATCTGCCTAAGCACATTTTATATA 1438

551 rgLysLysLeuGlyPhe 556
    |||||||||||||||
1439 GGCAGAAGGAACAATTC 1455
```

```
1 MetCysGlyLeuLeuGlyThrAlaAsnGlyAsnAlaGluAla.. 16
    ||||||||||  ::||:::||||  :::::::|||
94 ATGTGTGGCCTACTTGCTGGTTGCTCTGATGATTCACAAGCTAA 143

17 ......PheValProAlaLeuGluArgAlaLeuProCysMetArgHisA 31
        :::           ||:::::||       ::|||
144 ACGAGTTCGCATACTCGAGCTTCTCGCAGATTG........AAGCACC 184

31 rgGlyProAspAlaProAlaGlyThrTrpHisAspAlaAspAlaAlaPheGly 47
    ||||||||||  :::|||  ::::|||
185 GTGGGCCAGACTGGAGTGGGCTCCACCAACATGGTGATAACTATTTGGCT 234

48 PheAsnArgLeuSerIleIleAspIleAlaHisSerHisGlnProLeuAr 64
    ::::::||||||  ::::||  |||||||
235 CATCAAAGGTTAGCCATTGTTGATCCTGCTTCTGCTGATCAACCTCTC.. 282

64 gTrpGlyProAlaAspGluProAspArgTyrAlaMetThrPheAsnGlyG 81
        ::|||             :::||  |||||||
283 .......TTCAATGAAGACAAATCAATTATTGTCACGGTGAATGGAG 322

81 luIleTyrAsnTyrValGluLeuArgLysGluLeuSerAspLeuGlyTyr 97
    |||||||||:::  ||||||||||||||  :::  ::||
323 AAATCTACAATCATGAAGAGCTCAGAAAACAATTGCCCAAT......CAC 366

98 ThrPheAsnThrSerGlyAspGlyGluProIleValValGlyPheHisHi 114
    ||||||||||||||||||  :::
367 AAGTTTTTTACACAATGTGACTGTGATGTTATTGCACACCTGTACGAGGA 416

114 sTrpGlyGluSerValValGluHisLeuArgGlyMetPheGlyIleAlaI 131
    ::|||||||||||| |||::|||      ||||||||  :::||||
417 ACATGGAGAAAATTTTGTGGATATGTTAGACGTATATTTCGTTTGTTC 466

131 leTrpAspThrArgLysLysSerLeuPheLeuAlaArgAspAspAsnPheGly 147
    ::  |||||||||||||::::||  :::||||||||||  ::|||
467 TGCTGGATACTCGTGACAACAGTTTCATAGTTGCGAGGATGCTATAGGT 516

148 IleLysProLeuPheTyrAlaThrThrGluHisGlyThrValPheSerSe 164
    :::  ||  ||||||||  ||||||  :::||||||||
517 GTTACTTCCTTGTACATTGGTTGGGGACTAGATGGTTCTCTTGG..... 561

164 rGluLysLysThrIleLeuGluLeuMetAsnLeuAspLeuG 181
    ::||||||||:::  |||||||||:::||:::
562 ...............ATTGCATCAGAATTGAA........G 580

181 lyLeuAspLysArgThrIleGluHisTyrValAspLeuGlnTyrValPro 197
    ||||||::::  :::||||  ||||||||||  |||
581 GACTGAATGATGAA..TGTGAACATTTC.........GAAGTTTTT 615

198 GluProAspThrLeuHisIleAlaGlnIleSerArgLeuGluSerGlyCysTh 214
    |||                  :::  ||||||||||
616 CCGCCCGGTCACTTATACTCGAGCAAAGAAGA.............. 648

214 rAlaThrValArgProGlyLyLysGluGlnLysArgTyrPheLysSP 231
    |||                    :::::||||||||||::::::|
649 ..............GAGTTTCGTCGATGGTATAAATC 670

231 roGlnPhe..........ProValGlnLysVal 238
    ||:::             |||
671 CTCCATGGTTCAATGAGGCTAATATTCCGTCAACACCTTATGATCCTCTA 720

239 ValLysGlyGluGlnAspLeuPheAspArgIleAlaGlnValLeuGl 255
    ||||  :::::||||||||::::||||
721 GTT...............TTGAGGAACGCGCTTTGA 740

255 uAspSerValGlyLysHisMetArgAlaAspValThrValGlySerPheL 272
    |  :::||  ::|||||||  ||||     ||| ||:::
741 GAAGGCTGTGATAAAGAGGTTGATGACCGATGTGCCTTTCGGGGTTTTAC 790

272 euSerGlyIleIleAspSerThrAlaIleAlaProLeu.......... 284
    ||||||||||||||::::  ::  :::
791 TATCGGGAGGTTTGGATTCATCGTTGGTCGCGTCGTCACTGCTAGATAC 840
```

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-737-825-1

seq_documentation_block:
; Sequence 1, Application US/08737825
; Patent No. 5871922
; GENERAL INFORMATION:
; APPLICANT: SALMOND, GEORGE PEACOCK COPELAND
; APPLICANT: MCGOWAN, SIMON JAMES
; APPLICANT: SEBAIHIA, MOHAMMED
; APPLICANT: COX, ANTHONY RICHARD JOHN
; APPLICANT: HOLDEN, MATTHEW THOMAS GEOFFREY
; APPLICANT: PORTER, LAUREN ELIZABETH
; APPLICANT: BYCROFT, BARRIE WALSHAM
; APPLICANT: WILLIAMS, PAUL
; APPLICANT: STEWART, GORDON SIDNEY ANDERSON BIRNIE
; TITLE OF INVENTION: GENES INVOLVED IN THE BIOSYNTHETIC PATHWAY OF CARBAPENEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
;   STREET: P.O. Box 747
;   CITY: Falls Church
;   STATE: Virginia
;   COUNTRY: USA
;   ZIP: 22040-0747
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: MICROSOFT WORD97
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/737,825
;   FILING DATE: 03-JUN-1997
;   CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
;   NAME: SVENSSON, LEONARD R.
;   REGISTRATION NUMBER: 30,330
;   REFERENCE/DOCKET NUMBER: 1009-0105P
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703) 205-8000
;   TELEFAX: (703) 205-8050
;   TELEX: 248345
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 8341 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
;   ORGANISM: Erwinia carotovora
08-737-825-1

alignment_scores:
          Quality: 184.00   Length: 317
          Ratio: 1.243   Gaps: 11
Percent Similarity: 46.688   Percent Identity: 23.975

alignment_block:
US-09-786-474-2 x US-08-737-825-1   .

Align seg 1/1 to: US-08-737-825-1 from: 1 to: 8341

 246 LeuPheAspArgIleAlaGlnVal...LeuGluGluSerValGlyIleAspS 261
     ||| |||||| :|| ||| ::
2267 TTGATTGATCGCTACCTTAATGCTCCGCTTGAGGATTTAGGCGCCG...... 2311

 261 sMetArgAlaAspValThrValGlyTyrValGlySerPheLeuSerPheGlyIleGlyIleAspS 278
     ||| |||||||| ||||||||||||||||||::|||
2312 ...CGCTTTGAT...ACCGTAGGAATTCCCTTGTCAGGCGGTCTGGATT 2354

 278 erThrAlaIleAlaProLeuAlaLysArgHisAsnProAspLeuLeuThr 294

alignment continuation (right column):

                    ||::..:::: ||||||| ||| |||
2355 CCAGCCTGCTAACGCGCTTCGCCAGTCGTCATTCAAAAAATTGAATACG 2404
 295 PheThrThrGlyPheGluArgGluGlyLeuGlyTyrSerGluvalAspvalAlaAlaI 311
     ::||| ||| ||: ||| ::||
2405 TATTCGATTGGTACGGAACTCAGC....AATGAGTTTGAGTTTTCTCA 2448
 311 aGluSerAlaAlaAlaGlyAlaAlaGluHisIleValLysIleValSerP 328
     || ||| || |||||::||||||| |||::|||
2449 ACAGGTTGCTGATGCACTCGGTACACATCAGATGAAAATTCGTCCG 2498
 328 roGluGluTyrAlaAsnAlaIleProLysIleMetTrpTyr......Leu 342
     ||| ||| ||||::||| ||||||:::|||
2499 AAACTGAGTGATCAACGGCATCGAATCCATCTATTACAACGAAATA 2548
 343 AspAspProValAlaAspProSerLeuValProLeuTyrPheValAlaAlaI 359
     ||| ::||::: ||| |||
2549 TTTGACGGCTTATCCGCTGAAATCCAATCCGGGTTGTTCAATGTCTATCG 2598
 359 aGluAlaArgLysHisValLysValLeuSerGlyGluGlyvalAlaAspG 376
     ::|||:::   :::|||||::|| |||::||:::|||
2599 TCAGGCTCAGGGGCAGGTGTCTTCCATGCTCACCGGATATGGTTCCGACC 2648
 376 luLeuPheGlyGlyTyrThrIleTyrLysGluProLeuSerLeuAlaPro 392
     ||||||||||||
2649 TGCTCTTTGGCGGC................................. 2662
 393 PheGluLysIleProSerProLeuArgLysGlyLeuGlyLysLeuSerLy 409
     2662
 409 sValLeuProAspGlyMetLysGlyLysLysSerLeuLeuGluArgGlySerM 426
                            .ATACTGAAACCAGGAGCGC 2681
                            2663
 426 etThrMetGluGluArgTyrTyrGlyAsnAlaArgSerPheAsnPheGlu 442
     || ||| TATGACAATCCGAATCAGCTGCTTGCCGAG 2713
 443 GlnMetGlnArgValIleProTrpAlaLysArgGluTrpAspHisArgG1 459
     ||:::|| :::||| CGTTGGACAGGGGAGTT 2745
2714 CAAGTGTACCGGACG.................................. 2714
 459 uValThr.....AlaProIleTyrAlaGlnSerArgAsnPheAspProV 474
     ||| ||| TGCTACCCACGGTGCTTCCTGTTAC............. 2770
 474 aIAlaArgMetGlnHisLeuAspLeuPheThrTrpMetArgGlyAspIle 490
     2770
 491 LeuValLysAlaAspLysIleAsnMetAlaAsnSerLeuLeuArgVa 507
                        ||||||| GGCATTGATATTCGCCA 2787
 507 lProPheLeuAspLysGluValPheLysValAlaGluThrIleProTyrA 524
     |||||:: :: |||||||| |||:::
2788 CCCCTTCTCGGAGCCATTCCTAATCTCTATGTCATGCGCTACATCCTG 2837
 524 spLeuLysIleAlaAsnGlyThrThrLysTyrAlaLeuArgArgAlaLeu 540
     |||||| :::||| ||||||||::
2838 ATTACAAAATTTCGACAACGAAGTCAAAAACATCCTGCGTGAATACGCC 2887
 541 Glu......GlnIleValProProHisValLeuHisArgGlyLysLeuGl 555
     ::: || ||||||||| ||||||||::::||
2888 GATTCGCTGCAATTGCTGCCGAAAGACATTGTCTGGCGCCAAGAAAATCGG 2937

 555 y 555
     |
2938 C 2938

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-385-028-16

```
seq_documentation_block:
; Sequence 16, Application US/09385028
; Patent No. 6232106
; GENERAL INFORMATION:
;   APPLICANT:  Susan E. Jensen
;   APPLICANT:  Kwamena A Aidoo
;   APPLICANT:  Ashish S. Paradkar
;   TITLE OF INVENTION:  DNA Sequence Encoding Enzymes of Clavulanic
; Patent No. 6232106
;   TITLE OF INVENTION:  Acid Biosynthesis
;   NUMBER OF SEQUENCES:  25
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  JACOBSON, PRICE, HOLMAN & STERN, PPLC
;     STREET:  The Jenifer Building, 400 Seventh Street, N.W.
;     CITY:  Washington
;     STATE:  D.C.
;     COUNTRY:  U.S.A.
;     ZIP:  20004
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/09/385,028
;     FILING DATE:
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/790,462
;     FILING DATE:  29-JAN-1997
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  D. Douglas Price
;     REGISTRATION NUMBER:  24,514
;     REFERENCE/DOCKET NUMBER:  1418/P57452US2
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (202) 638-6666
;     TELEFAX:  (202) 39305350
;     TELEX:  RCA 248593 IDEA UR
; INFORMATION FOR SEQ ID NO:  16:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  1542 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-09-385-028-16

alignment_scores:
    Quality:   165.00          Length:   567
    Ratio:     0.693           Gaps:      23
Percent Similarity:  41.975    Percent Identity:  21.869

alignment_block:
US-09-786-474-2 x US-09-385-028-16    ..

Align seg 1/1  to: US-09-385-028-16   from:  1  to: 1542
```

```
 18 ValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyPro.. 33
    ||||||::::||::::                |||         |||||
 33 GTTCCTGGCCTCCGC.........CCGAACAGGCGGGGGCGGCCCCCCG 73

 34 ........AspAspAlaGlyThrTrpHis.AspAlaAspAlaAlaPhe 46
             |||  ||  |||  |||  |||:
 74 GCCCGGTCTTCGCGACACCGGGGCAGCCACCGACATCGACACGCCCCAG 123

 47 GlyPheAsnArgLeuSerIleIleAspIleAlaHisSerHisGlnProLe 63
    |||  ::  :||  ||::::  :::::||||:::
124 GGGGAGCGGCTCGCTCCGCG..GCGACCCTCGGTGCACGCC......... 159

 63 uArgTrpGlyProAlaAspGluProAspArg.................. 73
    |||:::||||
160 .....CCCTCGGTCGCGCCCGACCCGGCGCCGTGGCGCGCCTCCCTCA 199
```

```
 74 .........TyrAlaMetThrPheAsnGlyGluIleTyrAsnTyrVal 86
                :::::|||||||||||:::
200 CCGGCGCGCCCACCACCCGGTGCTCCGCGTGAGATCTACAACCGGGAC 249

 87 GluLeuArgLysGluLeuSerAspLeuGlyTyrThrPheAsnThrSerGl 103
    |||||              ::|||
250 GAACTC.........CTCTCCGTGCTGCCCGCGACCCGCCCGAGGG 290

103 yAspGlyGluProIleValValGlyPheHisHisTrpGlyGluSerValV 120
    ||||::|||         :::::::|||         ::
291 GGACCGGGAGCTGGTCCTGGCGCTGCTGGAACGCTATGACCTGCATGCCT 340

120 alGluHisLeuArgGlyMetPheGlyIleAlaIleTrpAspThrLysGlu 136
         :::::::|||  :::||||::::                ||
341 TCCGGCTGGTGAACGGGCCTTCGCGACCGTGTG...CGGACCGGGGAC 387

137 LysSerLeuPheLeuAlaArgAspGlnPheGlyIleLeuLysProLeuPheTy 153
    :::  |||  |||  |||              ||||||
388 CGGGTCCTG..CTCGCCACCGACCACCGCGTTCGGTGCCCGCTGTACAC 434

153 rAlaThrThrGluHisGlyThrValPheSerSerGluLysLysThrIleL 170
    :::::::||||           |||:|||  |||::::
435 CTGTGTGGCGCGGGCGAGGTCCGGCGTCCACCGAGGCCAAGGCGCTCG 484

170 euGluMetAlaGluGluMetAsnLeuAspLeuGlyLeuAspLysArgThr 186
                  :::::              |||  |||
485 CCCGCGCACCGGGACCCGGAAGGGTTCCCGCTCCGG...GACGCCCGCCGG 531

187 IleGluHisTyrValAspLeuGlnTyrValProGluProAspThrLeuH1 203
    :::              :|||  |||  ::
532 GTCGCCGGTCTGACCGGTGTCTACCAGGTGCCCGGGCGCCGTGATGGA 581

203 sAlaGlnIleSerArgLeuGluSerGlyCysThrAlaThrValArg..... 218
    :::  ||||||  :::||||  |||        |||
582 CATCGAC.........CTCGGCACGGCACCGCGTCACCCACCGCCACCT 622

219 .....ProGlyLysLeuGluGlnLysGlyArgTyrPheLysProGlnPhe 233
         ||||||              :::::
623 GGACCCCGGGC.........CTCTCCCCGCCATCCTGCCGGAGGGC 660

234 ProValGlnLysValValLysGlyLysGluGlnLeuPheAspAspArgIl 250
    |||:::||||||:::::
661 GAGGCCGTCGCGGCCGTGCGCGGCC..................... 684

250 eAlaGlnValLeuGluAspSerValGluLysHisMetArgAlaAspValT 267
    :::::::|||||
685 ........GCGCTGGAGAAGGCCGTGCCCCAGCGGGTCA 715

267 hrValGlySer......PheLeuSerGlyGlyIleAspSerThr 279
    ||  |||:::        ||||||||||||:::
716 CCCCCGGCGACACCCCGGTTGGTGTCTCCGGGAATCGACTCCTCCC 765

280 AlaIleAlaProLeuAlaLysArgHisAsnProAspLeuLeuThrPheTh 296
    ::::||  |||              :::|||  ||:
766 GGGGTCGCGGCCCTGTCGCGCACCGGGCGGGGAACTGGACACGGTGTC 815

296 rThrGlyPheGlu...........ArgGluGlyTyrSerGluValA 308
    ::  ::                  ||||||  :::  ||||
816 CATGGGACCGGACACGTCCAACGAGTTCCGCGAGGGCCCGGGGTCGTCG 865

308 spValAlaAlaGluSerAlaAlaAlaAlaIleGlyAlaGluHisIleValLys 324
                       ::::|||
866 AC.........CATCTTGGCCACCCGGACGGGAGATC 894

325 IleValSerProGluGlyTyrAlaAsnAlaIleProLysIleMetTrpTy 341
           ::||           :||:  ::::||
895 ACCATCCCGACCACCGAGCTGCTGGGCCAGCTCCGTACGCGGTGTCGGGC 944

341 rLeuAsp.....AspProValAlaAspProSerLeuProLeuTyrP 356
    ::  ||||                ||||:::::||||||
945 CTCCGAGTCGGTGACCCGGACCATCATCGAGTACCTGCTCCTGACAG 994
```

us-09-786-474-2.p2n.rni

```
356  heValAlaAlaGluAlaArgLysHisValLysValValLeuSerGlyGlu  372
995  CGCTCTACCGGGCGCTCGACGGGCCGGAGCGGCCGCCGCATCCTCACCGGGTAC  1044
                                  :::   :::||||:::||||

373  GlyAlaAspGluLeuPheGlyGlyTyrThrIleTyrLysGluProLeuSe  389
1045 GGCGCCGGACATCCCCCTCGGGGGC.....ATGCACCCCGAG.......  1080
     ||||||||           ::::::::::||

389  rLeuAlaProPheGluLysIleProSerProLeuArgLysGlyLeuGlyL  406
1081 .........GACCGGCTGCCCGCG.....................  1095
             :::::::::|||||:

406  ysLeuSerLysValLeuProAspGlyMetLysGlyLysSerLeuLeuGlu  422
1096 .CTGGACACCGTTCTCGCGCACGACATG.  1122
     |||::| ||||||          |||

423  ArgGlySerMetThrMetGluGluArgTyrTyrGlyAsnAlaArgSerPh  439
1123 ....................GCCACCTTCGA  1133
                        |||

439  eAsnPheGluGlnMetGlnArgValIleProTrpAlaLysArgGluTrpA  456
1134 CGGGCTGAACGAGATGTCCCCGTGCTGTCCACGCTGGCGGGCACTGGA  1183
     :::::::::::::|||::|    |||::|        ::::|||

456  spHisArgGluValThrAlaProIleTyrAlaGlnSerArgArgAsnPheAsp  472
1184 CCACCCAC.................................  1191
     :::

1184 CCACCCAC.  1191

473  ProValAlaArgMetGlnHisLeuAspLeuPheThrTrpMetArgGlyAs  489
1192 .....CCGTACTGGGACCGGAGGTCCTCGATCTGCTCGTCTGCGTCTGCGTCTGGAG  1236
          |||::||||||:::    :::::|

1191 ....................................................  1191

489  pIleLeuValLysAlaAspLysIleAsnMetAlaAsnSerLeuGluLeuA  506
1191 ....................................................  1191

506  rgValProPheLeuAspLysGluValPheLysValAlaAlaGluThrIlePro  522
1237 GCCGGGCTCAAGCGCGGCGGCGGCAACGGCGGACAAGTGGGTGCTGCGCCGCGC  1286

523  TyrAspLeuLysIleAlaAsnGlyThrThrLysTyrAlaLeuArgArgAl  539
1237 GCCGGGCTCAAGCGCGGCGGCGGCAACGGCGGACAAGTGGGTGCTGCGCCGCGC  1286

539  aLeuGluGlnIleValProHisValLeuHisIleArgLysLeuGly  555
1287 GATGGGCCGACGCCCTCCCGGCGGAGACCGTCAACCGCCCAAGCTGGGC  1335
     |::: ::||| :::::|||::::::|||| |||||||||
```

seq_name: /cgn2_6/ptodata/l/lna/6B_COMB.seq:US-09-385-028-13

documentation_block:
quence 13, Application US/09385028
Patent No. 6232106

GENERAL INFORMATION:
APPLICANT: Susan E. Jensen
APPLICANT: Kwamena A Aidoo
APPLICANT: Ashish S. Paradkar
TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic

Patent No. 6232106
TITLE OF INVENTION: Acid Biosynthesis
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PPLC
STREET: The Jenifer Building, 400 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

```
alignment_scores:
        Quality:   165.00            Length:    567
        Ratio:       0.693             Gaps:      23
Percent Similarity: 41.975   Percent Identity: 21.869

alignment_block:
US-09-786-474-2 x US-09-385-028-13

Align seg 1/1   to: US-09-385-028-13   from: 1   to: 11604

18   ValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyPro.. 33
     ||||||::::||:::                         ||   ||||||
1940 GTTCCTGGCCTCCGC.........CCGAACGGGCGGGGGGCCGGGCCCCCG 1980

34   .........AspAspAlaGlyThrTrpHis.AspAlaAspAlaAlaPhe  46
              ||| |||   ||||| |||   |||:::
1981 GCCCGGTCTTCGCACCCCGGGCAGCCACCGACATCGACACGCCCCCAG 2030

47   GlyPheAsnArgLeuSerIleIleAspIleAlaHisSerHisGlnProLe  63
     ||| |||  :::|||:||::||                           
2031 GGGGAGGCGTCGCTCGCG...GCGACCCTGGTGCACGCC.......... 2066

63   uArgTrpGlyProAlaAspGluProAspArg.................. 73
     |||::|| ||||||||
2067 .........CCCTCGGTCGCGCCGCACCGACGCCGGTGGCGCGCTCCCTCA 2106

74   .........TyrAlaMetThrPheAsnGlyGluIleTyrAsnTyrVal  86
              :::::|| |||||||||||||||||||| ||||
2107 CCGGCGCGCCACCACCGGTGCTGCGCCGGTGAGATCTACAACCGGGAC 2156

87   GluLeuArgLysGluLeuSerAspLeuGlyTyrThrPheAsnThrSerGl 103
     ||||||     :::::|||     |||||| |||     ::||
2157 GAACTC........CTCTCCGTGCTGCCCGCCGGACCCGCCGGAGGG 2197

103  YAspGlyGluProIleValValGlyPheHisHisTrpHisGlyGluSerValV 120
     |||:::|||  :::::|   :::::                :::
2198 GGACGGGAGCTGGTCTCCGGCGTCTGCTGGAACGCTATGACCTGCATGCCT 2247

120  alGluHisLeuArgGlyMetPheGlyIleAlaIleTrpAspThrLysGlu 136
     :::::::||||| |::::      :::::::      |||
2248 TCCGGCTGGTGAACGGGCCTTCGCCACCGTGGG...CGGACCGGGGAC 2294

137  LysSerLeuPheLeuAlaArgAspGlnPheGlyIleLeuLysProLeuPheTy 153
     ::: ||| |||||| :::::     :::     |||||||||||||
2295 CGGGTCCTG...CTCGCCACCGACCACCGCCGGTTCGGTGCCGGCTGTACAC 2341

153  rAlaThrThrGluHisGlyThrValPheSerGluLysLysThrIleL 170
     :::                             ||||:::|||
```

2342 CTGTGTGGCCGGGCGGAGGTCCGGCGTCCACCGAGGCCAAGGCCTCG 2391
170 euGluMetAlaGluGluMetAsnLeuAspLeuGlyLeuAspLysArgThr 186
         ::: |||||| |||::: ||| |||
2392 CCGCGCACCCGACCCGAAGGGCTTCCCGTCGCG...GACGCCCGCCGG 2438
187 IleGluHisTyrValAspLeuGlnTyrValProGluProAspThrLeuHi 203
         ::: |||||:   :::|||
2439 GTCGCGCAGTCTCACCGGTGTCTACCAGGTGCCCGCGGGCCCGTGATGGA 2488
203 sAlaGlnIleSerArgGluSerGlyCysThrAlaThrValArg..... 218
         ||| ||||| |||
2489 CATCGAC..........CTCGGCTCGGGCCACCCGTCACCCACCGCACCT 2529
219 ....ProGlyGlyLysLeuGluGlnLysArgTyrPheLysProGlnPhe 233
         ||||| ::||| |||:::
2530 GGACCCCGGGC............CTCTCCCCGCATCCTGCCGGAGGGC 2567
234 ProValGlnLysValLysValLysGluGlnAsnAspPheAspArgIl 250
         ::: |||||:::::
2568 GAGGCCGTCGGGCCCGTGCCGGCC.......................... 2591
250 eAlaGlnValLeuGluAspSerValGluLysHisMetArgAlaAspValT 267
         ||||| |||||
2592 ...........GGGCTGAGAGGCCCGTCGCCCAGCGGGTCA 2622
267 hrValGlySer............PheLeuSerGlyGlyIleAspSerThr 279
         || |||||| ||||||||||||
2623 CCCCCGGCGACACCCCGTTGGTGGTGCTCTCCGGGAATCGACTTCCTCC 2672
280 AlaIleAlaProLeuLeuAlaLysArgHisAsnProAspLeuThrPheTh 296
         :::|||||| ||| ||| ::|||
2673 GGGGTCGCGGCCTGTCGCCACCGGGCGACCTGGACACGGTGTC 2722
296 rThrGlyPheGlu.........ArgGluGlyTyrSerGluValA 308
         ||||||||| ||||
2723 CATGGGCACCGACACGTCCAACGAGTTCCGGGAGCCCGGGGTCGTCG 2772
308 spValAlaSerAlaAlaAlaAlaIleGlyAlaAlaHisIleValLys 324
         || |||||:::
2773 AC............CATCTGCGCACCCGACCACCGGGAGATC 2801
325 IleValSerProGluGluTyrAlaAsnAlaIleProLysIleMetTrpTy 341
         ::: ||| ::|||
2802 ACCATCCCGACCACCGAGCTGCTGGCCAGCTCCCGTACGCGGTGTGGGC 2851
341 rLeuAsp......AspProValAlaAspProSerLeuValProLeuTyrP 356
         ||| :::|||||||:::||
2852 CTCCGAGTCGGTGGACCCGGACATCATCGAGTACCGTCTCCCCCTGACAG 2901
356 heValAlaAlaGlyAlaGlyAlaArgGlyHisGlyAlaArgLeuSerGlyGlu 372
         ::: ||||||||| |||
2902 CGCTCTACCGGGCGCTCGACGGGCCCGAGGCGCGCATCCTCACCGGGTAC 2951
373 GlyAlaAspGluLeuPheGlyGlyTyrThrIleTyrLysGluProLeuSe 389
         ||||||| ::||||:::||
2952 GGCGCGGACATCCCCCTCGGGGGC......ATGGCACCGCGAG 2987
389 rLeuAlaProPheGluLysIleProSerProLeuArgLysGlyLeuGlyL 406
         ::::||||:::|||
2988 ...........GACCGGCTGCCCCGCG............. 3002
406 ysLeuSerLysValLeuProAspGlyMetLysGlyLysSerLeuLeuGlu 422
         ||||| |||||
3003 ..CTGGACACCGTTCTCGCCGCCACGACATG.................. 3029
423 ArgGlySerMetThrMetGluGluArgTyrArgGlyAsnAlaArgSerPh 439
         :::
3030 ....GCCACCTTCGA 3040
439 eAsnPheGluGlnMetGlnArgValIleProTrpAlaLysArgAlaTrpA 456
         ::|||||| |||:::
3041 CGGGCTGAACGAGATGTCCCCGTGTGTCCCACGCGTGGCGGGCACTGGA 3090

456 spHisArgGluValThrAlaProIleTyrAlaGlnSerArgAsnPheAsp 472
         ::|
3091 CCACCCAC................................ 3098
473 ProValAlaArgMetGlnHisLeuAspPheThrTrpMetArgGlyAs 489
3098 3098
489 pIleLeuValLysAlaAspLysIleAsnMetAlaAsnSerLeuGluLeuA 506
3098 3098
506 rgValProPheLeuAspLysGluValPheLysValAlaAlaGluThrIlePro 522
         ::::|||||||||
3099 ....CCGTACTGGGACCCGGAGGTCCTCCGATCTGCTGCGTCCTGCTGGAG 3143
523 TyrAspLeuLysIleAlaAlaAsnGlyThrThrLysThrTyrAlaLeuArgArgAl 539
         ||||| ::|||| |||||:::|||||
3144 GCCGGGCTCAAGCGGCGACACGGCCGGACAAGTGGGTGCTGCGCGCGCGC 3193
539 aLeuGluGlnIleValProProHisValLeuHisArgLysLysLeuGlyLeuGly 555
         |::: :::   ::||| :::|||||| ||||||||
3194 GATGGCCGACGGCCCTCCCGGCGGAGACCGTCAACCGGCCCAAGCTGGGC 3242

seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-385-028-1

```
;   ORIGINAL SOURCE:
;   ORGANISM:   Streptomyces clavuligerus
US-09-385-028-1

alignment_scores:
        Quality:      165.00              Length:        567
        Ratio:        0.693               Gaps:           23
Percent Similarity:   41.975       Percent Identity:   21.869

alignment_block:
US-09-786-474-2 x US-09-385-028-1

Align seg 1/1   to: US-09-385-028-1   from: 1  to: 15079


  18  ValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyPro..  33
      |||||||||:::||::|    |||  |||||   ||||||
3972  GTTCCTGCCCTCCGC.......CCGAACGGGCGGGGCGGCCGGGCCCCCG  4012

  34  .........AspAspAlaGlyThrTrpHis.AspAlaAspAlaAlaAlaPhe  46
              ::|||    ||||||||:::    ||::|    |||:::
4013  GCCCGGCGTCTTCGCGACCCGGGCAGCCACACCGACACGCCCCAG  4062

  47  GlyPheAsnArgLeuSerIleIleAspIleAlaHisSerHisGlnProLe   63
      |||    ||:::|    ||:::|||::||
4063  GGGGAGCCGTCGCTCGCG...GCGACCCTGGTGCACCGC..........  4098

  63  uArgTrpGlyProAlaAspGluProAspArg...................  73
      ||||||||
4099  .....CCCTCGGTCGCCGACCGGGCTGCGCGCTCCCTCA  4138

  74  .........TyrAlaMetThrPheAsnGlyGluIleTyrArgAsnTyrVal  86
              :::::|    ||||||||||||||
4139  CCGGCGGCCCCACCCGGGTGCTCGCCCGGTGAGATCTACAACCGGGAC  4188

  87  GluLeuArgLysGluLeuSerAspLeuGlyTyrThrPheAsnThrSerGl  103
      |||||||    |||||    |||    ::|
4189  GAACTC...CTCTCCGTCGCTGCCCGGACCCGGCGGAGGG  4229

 103  yAspGlyGluProIleValValGlyPheHisHisTrpGlyGluSerValV  120
      ||||:::||||    :::::||
4230  GGACGGGGAGCCGGCTGCTCCGGACCTATGCAACCCTCATGCCT  4279

 120  alGluHisLeuArgGlyMetPheGlyIleAlaIleTrpAspThrLysGlu  136
      :::    ||||    ||||||||:    ||::|||
4280  TCCGGCGTGGTGAACGGGCGTTCGCGACCGGTG...CGGACCGGGGAC  4326

 137  LysSerLeuPheLeuAlaArgArgAspGlnPheGlyIleLeuLysProLeuPheTy  153
      :::    |||||||||    |||    ||:::    ||::||
4327  CGGGTCCTG...CTCGGCCACCGGCACCGGTTCGGTGCCCGCTGTACAC  4373

 153  rAlaThrThrGluHisGlyThrValPheSerGlyGluLysLysThrIleL  170
      :::    ||::|||    ::|::|||::||||
4374  CTGTGTGGGCCGGCCGGAGGTCCGGGCGTCCACCGAGCCAAGGCGTCG  4423

 170  euGluMetAlaGluGluMetAsnLeuAsnLeuGlyLeuAspLysLysArgThr  186
      |||    |||||||    ||||    ||||||
4424  CCGGCGCACCGCGACCCCGAAGGGCTTCCCGCTCGCG...GACGCCCGGCCGG  4470

 187  IleGluHisTyrValAspLeuGlnTyrValProGluProAspThrLeuHi  203
      :::    ||||::|    |||||    |||
4471  GTGCCGGCGTGACGGTCTACCAGGTGCCCGGGCGCCGTGATGGA  4520

 203  sAlaGlnIleSerArgGluSerGlyCysThrAlaThrValArg.......  218
      :::    |||    ||||||||||
4521  CATCGAC.......CTCGGGCTGGGCACCCGGGTCACCCACCGCACCT  4561

 219  ....ProGlyGlyLysGluGlnLysLysLeuArgThrPheLysProGlnPhe  233
      ||||||    |||::|    :::    ||:::
4562  GGACCCCCGGGC..........CTCTCCCGCCATCCTGCCGAGGGGC  4599

 234  ProValGlnLysValValLysGlyLysGluGlnAspLeuPheAspArgIl  250
      :::    ::::::|||::
```

```
4623  GAGGCCGTCCGGCCGTGCGGCC..................  4623

 250  eAlaGlnIleValLeuGluAspSerValGluLysHisMetArgAlaAspValT  267
      :::    ||::|
4624  .........GCGCTGGAGGAGGCCGTCCGGCCCAGCGGGTCA  4654

 267  hrValGlySer.............PheLeuSerGlyGlyIleLeuAspSerThr  279
      ||    ||::::    ||||||||||||||||::
4655  CCCCCGGCGCACCCCGTTGGTGGTGCTCTCCGGCGGAATCGACTCCTCC  4704

 280  AlaIleAlaProLeuAlaLysArgHisAsnProAspLeuLeuThrPheTh  296
      :::    ::::|    ||    ::|||
4705  GGGGTCGCGCGCCTGTCGCCACCGGGCGGCCGGGAACTGACACGGTGTC  4754

 296  rThrGlyPheGlu.............ArgGluGlyTyrSerGluValA  308
      :::    |||    ||
4755  CATGGCACCGACACGTCCAACGAGTTCCGGCAGCCCGGGCGGTCGTCG  4804

 308  spValAlaAlaGluSerAlaAlaAlaAlaIleGlyAlaGluHisIleValLys  324
      |||    :::::|||
4805  AC..........CATCTGCGCACCCGGCCACCGGGAGATC  4833

 325  IleValSerProGluTyrAlaAsnAlaIleProLysIleMetTrpTy  341
      :::    ||||
4834  ACCATCCCGACCACCGAGCTGCTGCGCAGCTCCCTACCGCGTGTGGGC  4883

 341  rLeuAsp.......AspProValAlaAspProSerLeuValProLeuTyrP  356
      :::    |||||||
4884  CTCCGGACTCGGTGGACCCGGACCATCAGTAACCTGCTCCCCCTGACAG  4933

 356  heValAlaAlaGluAlaArgHisValLysValValLeuSerGlyGlu  372
      :::    :::|||::::|||
4934  CGCTCTACCGGCGCTCGACGGGCCGGAGCCCGCATCCTCACCGGGTAC  4983

 373  GlyAlaAspGluLeuPheGlyGlyTyrThrIleTyrLysGluProLeuSe  389
      :::    |||
4984  GGCGCGGACATCCCCCTCGGGGGC.......ATGCACCCGGAG.......  5019

 389  rLeuAlaProPheGluLysIleProSerProLeuAargLysGlyLeuGlyL  406
      :::    ::::||::||||
5020  .........GACCGGCTGCCCGCG..................  5034

 406  ysLeuSerLysValLeuProAspGlyMetLysGlyLysSerLeuLeuGlu  422
      |||::::    ||||||
5035  ..CTGGACACCGTTCTGCGCCACGACATG..................  5061

 423  ArgGlySerMetThrMetGluGluArgTyrTyrGlyAsnAlaAlaArgSerPh  439
      :::    |||||
5062  ..........................GCCACCTTCGA  5072

 439  eAsnPheGluGlnMetGlnArgValIleProTrpAlaLysArgGluTrpA  456
      :::    :::||::|||||
5073  CGGGCTGAACGAGATGTCCCCGGTGCTGTCCACGCTGGCGGGCACTGGA  5122

 456  spHisArgGluValThrAlaProIleTyrAlaGlnSerArgAsnPheAsp  472
      :::::
5123  CCACCCAC.......................................  5130

 473  ProValAlaArgMetGlnHisLeuAspLeuPheThrTrpMetArgGlyAs  489
      :::
5130  ......................................  5130

 489  pIleLeuValLysAlaAspLysIleAsnMetAlaAsnSerLeuGluLeuA  506
      :::
5130  ......................................  5130

 506  rgValProPheLeuAspLysGluValPheLysValAlaGluThrIlePro  522
      |||:::||||    :::||
5131  ....CCGTACTGGGACCCGGAGGTCCTCGATCTGCCGCTCGGAG  5175

 523  TyrAspLeuLysIleAlaAsnGlyThrThrLysTyrAlaLeuArgArgAl  539
      :::||||:::
5176  GCCCGGGCTCAAGCGGCGGCGACCGGACAAGTGGGTGCTGCCGCCCGGC  5225
```

```
 539 aLeuGluGlnIleValProProHisValLeuHisArgLysLeuGly 555
     |::: :::  ||  :::::::||| |||||||||| |||||||||
5226 GATGGCCGACCCCTCCCGCGGAGACCGTCAACCGCCCAAGCTGGGC 5274
```

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-599-171A-22

seq_documentation_block:
; Sequence 22, Application US/08599171A
; Patent NO. 5814473
;   GENERAL INFORMATION:
;   APPLICANT: WARREN, Patrick V.
;   TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES
;   NUMBER OF SEQUENCES: 32
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
;   ADDRESSEE: CECCHI, STEWART & OLSTEIN
;   STREET: 6 BECKER FARM ROAD
;   CITY: ROSELAND
;   STATE: NEW JERSEY
;   COUNTRY: USA
;   ZIP: 07068
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3.5 INCH DISKETTE
;   COMPUTER: IBM PS/2
;   OPERATING SYSTEM: MS-DOS
;   SOFTWARE: WORD PERFECT 5.1
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/599,171A
;   FILING DATE: Concurrently
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: HERRON, CHARLES J.
;   REGISTRATION NUMBER: 28,019
;   REFERENCE/DOCKET NUMBER: 331400-38
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 201-994-1700
;   TELEFAX: 201-994-1744
;   INFORMATION FOR SEQ ID NO: 22:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1779 NUCLEOTIDES
;   TYPE: NUCLEIC ACID
;   STRANDEDNESS: SINGLE
;   TOPOLOGY: LINEAR
;   MOLECULE TYPE: GENOMIC DNA
US-08-599-171A-22

alignment_scores:
              Quality: 154.50   Length: 580
              Ratio: 0.592   Gaps: 27
Percent Similarity: 45.000   Percent Identity: 20.000

alignment_block:
US-09-786-474-2 x US-08-599-171A-22

Align seq 1/1 to: US-08-599-171A-22   from: 1   to: 1779

  1 MetCysGlyLeuGlyIleIleLeuThrAlaAsnGlyAsnAlaGluAlaPh 17
    ||||||||| :::|| ::   ::::
  1 ATGTGCGGAGTAGTCGGATACGTA..........GGGAGGGATTTAGC 38

 17 eValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA 34
    ::||| ::: ||||| :::::::|||||| |
 39 CCTTCCTATAGTCCTCGAGCTCTTGAGAGACTCGAATACAGGGGTTACG 88

 34 spAspAlaGlyThr.................................. 38
    ||:::::
 89 ACTCCCGGAGTTGCCCTTATAGAAGACGGGAAACTCATAGTTGAAAAG 138
```

```
 39 .................................Trp.........H1 40
                                     |||
139 AAGAAGGGAAAGATAGGAACTCGTTAAAGCCTATGGGAAAGGATTA 188

 40 saspAlaAspAlaAlaPheGlyPheAsnArgLeuSer.......... 52
    ::: ||| ::::::||||| ||||
189 CAAGGCTAAAACGGGTATAGGTCACACGCCTGGCAACCCACGGAAAGC 238

 53 ..IleIleAspIleAlaHisSerHisGlnProLeuArgTrpGlyProAla 68
      ::: |||||| |||
239 CCACGGACGAGAACGCCACCCCCAC.................ACC 267

 69 AspGluProAspArgTyrAlaMetThrPheAsnGlyGluIleTyrAsnTy 85
    |||||  :::::::::|||   ||||  ||||
268 GACGAAAAGGTGAGTTTGCAGTAGTTCCACACGGATAATAGAAAACTA 317

 85 rValGluLeuArgLysGluLeuSerAspLeuGlyTyrThrPheAsnThrS 102
    |||||||||| :::::|||||||| |||||||||| |||
318 CTTAGAACTAAAAGAGAACTAAAGAAGAAGGAAGGTGTAAAGTTCAGGTCCG 367

102 erGlyAspGlyGluProIle......ValValGlyPheHisHisTrpGly 116
    :: ||| ||| |||               ||||||  :::::  ||
368 AAACAGACACAGAAGTTATAGCCCACCTCATAGCGAAGAACTACAGGGGG 417

117 Glu.........SerValValGluHisLeuArgGlyMetPh 127
    :::                        :::::::||| ||
418 GACTTACTGGAGGCGTTTAAAAACCGTAAAGAAATTAAGGGTGCTTT 467

127 eGlyIleAlaIleTrpAspThrArgLysGluLysSerLeuPheLeuAlaArgA 144
    |||||        ||:::||||||  :::::::|||||:::: ||:::
468 TGCCTTGCGGTTATAACGGTTCACGACCAAACAGACTAATA........ 510

144 spGlnPheGlyIleLeuLys........ProLeuPheTyrAlaThrThrGlu 157
    |||:::||                      |||    :::  |||
511 .GGAGTGAAGCAGGGAGTCCTTTAATCGTCGGACTCGGAGAA 552

158 HisGlyThrValPheSerGluLysLysThrIleLeuGluMetAlaGl 174
    :::                         :::::::|||||  |
553 GGAGAAAACTTCCTCGGCTTCAGATATTCCCGCAATACTTCCTTACACGAA 602

174 uGluMet.......................................... 176
    :::::::
603 AAAGATTATTGTTCTTGATGACGGGGAAATAGCGGACCTGACTCCCGACA 652

177 ..................AsnLeuAsp...................... 179
                      |||:::::
653 CTGTGAACATTTACAACTTTGAGGGAGCCCGTTCAAAGGAAGTAATG 702

180 ...............LeuGlyLeuAspLysArgThrIleGluHisTy 190
                   :::::            ::::::||:::|||
703 ATTACGGCCCTGGGATCTTGTTTCTGCGGAAAAGGTGGTTTAAACACTT 752

190 rValAspLeuGlnTyrValProGluPro..........AspThrLeuH 203
    :::                         ||||||||||
753 CATGCTAAAAGAGATATACGAACAGCCCAAAGCCATAAACGACACACTCA 802

203 lsAlaGlnIleSerArgLeuGluSer...................... 211
    ::: ::                 :::|||
803 AGGGTTCCTCTCAACCGAAGACGCCAATACCCTTAAGTAAAAGACTTC 852

212 ...............GlyCysThrAlaThrValArgProGly...... 220
                   |||||||   :::::::: ||
853 AGAAGGGTTTAATAATAGCGTGCGGGCACCTCTTACCACGCGGCGTTCGT 902

221 .GlyLysGlyLeuGlnLysArgTyrPheLysProGlnPheProValGlnL 237
     ||||||     ||||||  :::::::::||||||
903 CGGAAAGTTACTGCATAGAGAGATTT.........GCCAGTGTTCCCACAGAGG 946

237 ysVal............................ValLysGlyLysGluGln 244
    :::                              |||
947 TAATTTACGGCCTTCGGAATTCAGTTATGCGGACGTTCCCGTTTCGGACAAG 996
```

```
COMPUTER READABLE FORM:
    MEDIUM TYPE: Diskette
    COMPUTER: IBM Compatible
    OPERATING SYSTEM: Windows95
    SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/646,590B
    FILING DATE: 08-May-1996
    CLASSIFICATION: 435
PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 08/599,171
    FILING DATE: 09-FEB-1996
PRIOR APPLICATION DATA:
    APPLICATION NUMBER: PCT/US97/01094
    FILING DATE: 21-January-1997
    CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
    NAME: Haile, Ph.D., Lisa A.
    REGISTRATION NUMBER: 38,347
    REFERENCE/DOCKET NUMBER: 09010/017001
TELECOMMUNICATION INFORMATION:
    TELEPHONE: 619/678-5070
    TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
    LENGTH: 1779 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
FEATURE:
    NAME/KEY: Coding Sequence
    LOCATION: 1...1776
US-08-646-590B-22


alignment_scores:
    Quality:  154.50                      Length:     580
    Ratio:      0.592                      Gaps:        27
Percent Similarity: 45.000    Percent Identity:    20.000

alignment_block:
US-09-786-474-2 x US-08-646-590B-22

Align seg 1/1  to: US-08-646-590B-22   from: 1  to: 1779

  1 MetCysGlyLeuGlyLeuLeuThrArgAlaAsnGlyAsnAlaGluAlaPh 17
    ||||||||||||  :::: ||||          ::
  1 ATGTGCGGGATAGTCGGATACGTA................GGGAGGGATTTAGC 38

 17 eValProAlaLeuGlyLeuArgAlaLeuProCysMetArgHisArgGlyProA 34
    :::||| |||||  ||||||  |||||||
 39 CCTTCCTATAGTCCTCCGGAGCTCTTGAGAGACTCGAATACAGGGGTTACG 88

 34 spAspAlaGlyThr.................................... 38
    ||:::||||||:::
 89 ACTCCGCGGGAGTTGCCCTATAGAGACGGGAAACTCATAGTTGAAAAG 138

 39 ..............................................Trp.......H1 40
                                                  |||
139 AAGAAGGGAAAGATAAGGAACTCGTTAAAGCGCTATGGGAAAGGATTA 188

 40 sAspAlaAspAlaAlaPheGlyPheAsnArgLeuSer.............. 52
    ||| :::::::||| ::|||  :::
189 CAAGGCTAAAACGGTATAGGTCACACGCTGGCAACCCACCGAAAGC 238

 53 ..IleIleAspIleAlaHisSerHisGlnProLeuArgTrpGlyProAla 68
         ||::|||   |||||||||||  :|||||
239 CCACGGACGGAGAAGCGCCACCCCAC................ACC 267

 69 AspGluProAspArgTyrAlaMetThrPheAsnGlyIleLeuTyrAsnTy 85
    ||||| ||||||||  :::|  ||||| ||  ||||
268 GACGAAAAGGTGAGTTTCAGTAGTTCAACGGGATAATAGAAAACTA 317
```

```
245 AspLeuPheAspArgIleAlaGln................ 252
    |||:::|||  ||:::|||
997 GATATCGTTATCGAATTCCCAGTCAGGAGGACCGCTGACACAAAGTT 1046

253 .ValLeuGluAspSerValGluLysHisMetArgAspValThrValG 269
    :::::||| |||||| |||||  ::  ||||||
1047 TGCCCTTCAGTCCGCAAAGGAAAG.......GGAGCCCTTTACCGTGG 1087

269 ly.....SerPheLeuSerGlyLyIleAspSerThrAlaIleAlaPro 283
    ||  ::: :::::::::||||||||
1088 GACTCGTAAACGTAGTGGGAAGTGCCATAGACAGGGAGTCGGACTTTCC 1137

284 LeuAlaLysArgHisAsnProAspLeu............LeuThrPh 295
    ||| :: :::|||:::|||      :::||||
1138 CTTCACACACGTGGGACCCGAAATAGGCGTGCGGCTACAAAGACCTT 1187

295 eThrThrGlyPhe...........................G 300
    |||:::|||
1188 CACCGCACAGTTCACCGCACTCACGCCCTTTCGGTAAGGGAAAGTGAGG 1237

300 luArgGluGly.........TyrSerGluValAspSerValAlaAlaGlu 312
    ||||||||:::             ::::|||
1238 AGAGGGAAAATCTAATAGACTCCTTGAAAAGGTTCCATCACTCGTTGAA 1287

313 SerAlaAlaIleGlyAlaGluHisIleValLysIleValSerProGl 329
    :::::: ||||||||  ||:::|||
1288 CAAACACTGAACACC..GCAGAAGAGTGGAGAAGGTAGCGGCGAAAGTA 1334

329 uGluTyrAlaAsnAlaIleProLysIleMetTrpTyrLeuAspAspProV 346
    ||||||||||||||  :::||| ::: :::||||
1335 CATGAAAAAAGAAAAACATGCTTTACCTCGGAAGGTACTTAAATTACCCCA 1384

346 alAla.......AspProSerLeuValProLeuTyrPheValAlaAla 359
    :||| :::||| ::: :: ||| |||
1385 TAGCGCTGGAGGGAGCTCTTAAACTTAAAGAAATTTCTTACATACACGCG 1434

360 GluAla.............ArgLysHisValLysValValLeuSerGl 371
    ||:::             ||||||  :::|||
1435 GAAGGTTATCCCCAGGGAGATGAAGACGCTGTCCCATAGCCCTCATAGA 1484

371 yGluGlyAlaAspGluLeuPheGlyGlyTyrThrIleTyrLysGluProL 388
    |||:::::: ||||| ::|||||
1485 CGAAAACATGCCGGTTGTGGTA.................. 1506

388 euSerLeuAlaPro..........PheGluLysIleProSerProLeu 400
    ::||||||  ::|||||||||  :::
1507 ....ATCGGCACCGAAAGACAGGGTTTACGAGAGATACTCTCAAACGTA 1551

401 ArgLysGlyLeuGlyLysLeuSerLysValLeuProAspMetLysGl 417
    ::::::|| ||| ::::|||||
1552 GAAGAGGTTCTCGCAAGAAAGGGAAGGTTATTTCTGTAGCCTTAAAGG 1601

417 yLysSerLeuLeuGluArgGlySerMetThrMetGluGlu 430
    |||  :::   ||:::|| |||
1602 AGACGAAAACTCTCAAAAGCAAATCCGAGAGCGTTATGGAA 1641
```

```
seq_name: /cgn2_6/ptodata/1/1na/5B_COMB.seq:US-08-646-590B-22

seq_documentation_block:
; Sequence 22, Application US/08646590B
; Patent No. 5962283
GENERAL INFORMATION:
    APPLICANT: Warren, Patrick V.
    APPLICANT: Swanson, Ronald V.
    TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES
    NUMBER OF SEQUENCES: 42
    CORRESPONDENCE ADDRESS:
    ADDRESSEE: Fish & Richardson, P.C.
    STREET: 4225 Executive Square, Suite 1400
    CITY: La Jolla
    STATE: CA
    COUNTRY: US
    ZIP: 92037
```

```
 85 rValGluLeuArgLysGluLeuSerAspLeuGlyTyrThrPheAsnThrS  102
    :||:|||||||:::|||||||:::
318 CTTAGAACTAAAGAGGAAGAACTAAAGAAGGAGGAAAGTGTAAAGTTCAGGTCCG  367

102 erGlyAspGlyGluProIle.....ValValGlyPheHisHisTrpGly  116
    ::| ||| |||:::          |||:::: : ||||:
368 AACACAGACAGAAGTTATAGCCACCTCACTGCCGAAGAACTACAGGGGG  417

117 Glu...........SerValValGluHisLeuArgGlyMetPh  127
    ::|           :::::|||:||  :||:::|||  ||
418 GACTTACTGGAGGCCCGTTTAAAAACCGTAAAGAAATTAAAGGGTGCTTT  467

127 eGlyIleAlaIleTrpAspThrLysGluLysPheLeuLysPheLeuAlaArgA  144
    :::||:|||| |||  ||| :::::::
468 TGCCCTTGCCGTTATAACGGTCACGAACCAAACAGACTAATA.......  510

144 spGlnPheGlyIleLys........ProLeuPheTyrAlaThrThrGlu  157
    ||||:|        ||||||
511 ........GGAGTGAAGCAGGGAGTCCTTTAATCGTCGGACTCCGAGAA  552

158 HisGlyThrValPheSerGluLysThrIleLeuGluMetAlaGl  174
    ::: ::::|||:||||  :::||||||
553 GGAGAAAACTTCCTCGCTTCAGATATTCCCGCAATACTTCCTTACACGAA  602

174 uGluMet..........  176
    :::::::
603 AAAGATTATTTGTTCTTGATGACGGGGAAATAGCGGACCTGACTCCCGACA  652

177 .AsnLeuAsp.........  179
    ||| ||||
653 CTGTGAACAACATTTACAACTTTGAGGAGGAGCCCGTTTCAAAGGAAGTAATG  702

180 .........LeuGlyLeuAspLysArgThrIleGluHisTy  190
    :::::|:|||  :::|||:||
703 ATTACGCCCTGGGATCTTGTTTCTGCGGAAAAGGGTGGTTTAAACACTT  752

190 rValAspLeuGlnTyrValProGluPro..........AspThrLeuH  203
    :::|||  :::|||||||:::||||||
753 CATGCTCAAAAGAGATATACGAACAGCCCAAAGCCATAAACGACACACTCA  802

203 isAlaGlnIleSerArgLeuGluSer............  211
    ||:||:
803 AGGGTTTCCTTCACCGAGACGCAATACCCTTTAAGTTAAAAGACTTC  852

212 .............GlyCysThrAlaThrValArgProGly.....  220
    ||||| ::::::  |||
853 AGAAGGGTTTTAATAATAGGGCTGCGGACCTCTTACCACGCGGGCTTCGT  902

221 .GlyLysLeuGluGlnLysArgTyrPheLysProGlnPheProValGlnL  237
    ||||||        |||:::
903 CGGAAAGTACTGGATAGAGAGATTT.....GCAGGTGTTCCCACAGAGG  946

237 ysVal.............ValLysGlyLysGluGln  244
    :::|            |||
947 TAATTTACGCGTTCGGAATTCAGGTATGCGGACGTTCCCGTTTCGGACAAG  996

245 AspLeuPheAspArgIleAlaGln.............  252
    ||:||:||        |||:::|||
997 GATATCGTTATCGGAATTTCCCAGTCCCAGTGACGAGAGACCGCTGACAAAGTT  1046

253 .ValLeuGluAspSerValGluLysHisMetArgAlaAspValThrValG  269
    |||:::|||   :::||||||
1047 TGCCCTTCAGTCCGCAAAGGAAAG.......GGACCCTTACCGTGG  1087

269 ly.......SerPheLeuSerGlyGlyGlyIleAspSerThrAlaIleAlaPro  283
    ||||||        :::|||||||
1088 GACTCGTAAACGTAGTGGGAAGTGCCATAGACCAGGGAGTCGGACTTTTCC  1137

284 LeuAlaLysArgHisAsnProAspLeu...........LeuThrPh  295
    |||   :::|||
1138 CTTCACACACATGCGGGACCCGAAATAGGGCGTGGCGGCTACAAAGACCTT  1187
```

```
295 eThrThrGlyPhe..............G  300
    ||||:::|||  |||
1188 CACCGCACAGTTCACCGCACTCACCGCCCTTCCGTAAGGGAAGTGAGG  1237

300 luArgGluGly.........TyrSerGluValAlaSpValAlaAlaGlu  312
    ||||||||:::    :::|||
1238 AGAGGGAAAATCTAATAAGACTCCTTGAAAAGGTTCCATCACTCGTTGAA  1287

313 SerAlaAlaAlaIleGlyAlaGluHisLysIleValLysValLysSerProGl  329
    :::|||           |:::|||
1288 CAAACACTGAACACC..GCAGAAGAGTGGAGAAGGTAGCGGAAAAGTA  1334

329 uGluTyrAlaAsnAlaIleProLysIleMetTrpTyrLeuAspAspProV  346
    :::   ::: :::||||||:::
1335 CATGAAAAAGAAAAACATGCTTTACCTCGGAAGGTACTTAAATTACCCCA  1384

346 aIAla..........AspProSerLeuValProLeuTyrPheValAlaAla  359
    :::||       :::|||
1385 TAGCGCTGGAGGGAGCTCTAAACTTAAAGAAATTTCTACATACACGCG  1434

360 GluAla...........ArgLysHisValLysValValLeuSerGl  371
    |||:::         ||||||
1435 GAAGGTTATCCCGAGGGGGAGATGAAGCACGGTCCCATAGCCCTCATAGA  1484

371 yGluGlyAlaAspGluLeuPheGlyGlyTyrThrIleTyrLysGluProL  388
    ||::::
1485 CGAAAACATGCCCGGTTGTGTA..............  1506

388 euSerLeuAlaPro...........PheGluLysIleProSerProLeu  400
    :::||:::||    |||||||:::
1507 .....ATCGCACCGAAAGACAGGTTTACGAGAGATACTCTCAAACGTA  1551

401 ArgLysGlyLeuGlyLysLeuSerLysValLeuProAspGlyMetLysGl  417
    :::|||        :::||||||:::
1552 GAAGAGGTTCTCGCAAGAAAGGAAGGGTTATTTCTGTAGGCTTTAAAGG  1601

417 yLysSerLeuLeuGluLeuArgGlySerMetThrMetGluGlu  430
    ||:::|       ::::||
1602 AGACGAAAACTCTCAAAAGCAAATCCGAGAGCGTTATGGAA  1641
```

us-09-786-474-2.p2n.rni

```
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:   201-994-1700
;     TELEFAX:     201-994-1744
; INFORMATION FOR SEQ ID NO:   22:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  1779 NUCLEOTIDES
;     TYPE:  NUCLEIC ACID
;     STRANDEDNESS:  SINGLE
;     TOPOLOGY:  LINEAR
;   MOLECULE TYPE:  GENOMIC DNA
US-09-069-226-22

alignment_scores:
       Quality:  154.50        Length:   580
         Ratio:  0.592           Gaps:    27
Percent Similarity:  45.000    Percent Identity:  20.000

alignment_block:
US-09-786-474-2  x  US-09-069-226-22

ign seg 1/1  to: US-09-069-226-22  from: 1  to: 1779

  1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPh  17
    |||||||||||:::|||                .GGGAGGGATTTAGC  38
  1 ATGTGCGGGATAGTCGGATACGTA.................GGGAGGGATTTAGC

 17 eValProAlaLeuGluArgAlaLeuProCysMetArgHisArgGlyProA  34
    ::|||   ||||||          ::|||:|:|||||| |
 39 CCTTCCTATAGTCCTCGGAGCTCTTGAGAGACTCGAATACAGAGGGTTACG  88

 34 spAspAlaGlyThr...............................  38
    |||:|:||||:|||
 89 ACTCCGCGGGAGTTGCCCTTAGAGAGACGGGAAACTCATAGTTGAAAAG  138

 39 ...................................Trp.........H1  40
                                       |||        ||
139 AAGAGGGAAAGATAAAGGAACTCGTTAAAGCCGCTATGGGGAAAGGATTA  188

 40 saAspAlaAspAlaPheGlyAlaAsnArgLeuSer................  52
    |||   ::|||||||  ::|||:|
189 CAAGGCTAAAACGGGTATAGGTCACACACGTCTGGGCAACCCACGGAAAGC  238

 53 ..IleIleAspIleAlaHisSerHisGlnProLeuArgTrpGlyProAla  68
    :|:   |||:|:        |||               :::
239 CCACGGACGAGAACGCCCACCCCCAC...................ACC  267

 69 AspGluProAspArgTyrAlaMetThrPheAsnGlyGluIleTyrAsnTy  85
    ||||||   :::|||:|:|:   ||||||   |||   |||||
268 GACGAAAAAGGTGAGTTGCAGTAGTTCACAACGGGATAATAGAAAAACTA  317

 85 rValGluLeuArgLysGluLeuSerAspLeuGlyTyrThrPheAsnThrS  102
    :::|||||||:::|||:|:|||||            |||:|:::|||
318 CTTAGAACTAAAGAGGAACTAAAGAGGAAGGTGTAAGTTCAGGTCCG  367

102 erGlyArgGlyAspGluProIle......ValValGlyPheHisHisTrpGly  116
    ::::|||||| |||         |||  ||| |||:|:
368 AAACAGACACACGAAGTTATAGCCCAACTTCATAGCGAAGAACTACAGGGGG  417

117 Glu.............SerValValGluHisLeuLeuArgGlyMetPh  127
    :::              |||:|:::|||  ||
418 GACTTACTGAGGCCGTTTTAAAGAAACCCTAAAGAAATTAAAGGGTGCTTT  467

127 eGlyIleAlaIleTrpAspThrLysGluLysSerLeuPheLeuAlaArgA  144
    ::|||:|:|||| |||  ::::|||:|:||||||
468 TGCCTTTCGCGGTTATAACGGTTCAGACCAAACAGACTAATA.......  510

144 spGlnPheGlyIleLeuLys.........ProLeuPheTyrAlaThrThrGlu  157
    |||:|:|||  |||         :|:|||  |||
511 ........GGAGTGAAGCAGGGGAGTCCTTAATCGTCGGACTCGGAGAA  552

158 HisGlyThrValPheSerGluLysGluLysThrIleLeuGluMetAlaGl  174
    :::   ::::|||:|:|||||||
553 GGAGAAACTTCCTCGCTTCAGATATTCCGCAATACTTCCTTCACGAA  602

174 uGluMet............................................  176
    :::|||
603 AAAGATTATTGTTCTTGATGACGGGAGAAATAGCGGACCTGACTCCCGACA  652

177 ...........AsnLeuAsp...............................  179
               |||:|:|:
653 CTGTGAACATTTACAACTTTGAGGGAGGAGCCCGTTCAAAGGAGTAATG  702

180 ...........LeuGlyLeuAspLysArgThrIleGluHisTy  190
                              |||:  :|||:
703 ATTACGCCCTGGGATCTTGTTTCTCGGAAAAGGTGGTTTAAACACTT  752

190 rValAspLeuGlnTyrValProGluPro..........AspThrLeuH  203
    :::           |||:|:   |||||||||
753 CATGCTAAAAGAGATATACGAACAGCCCAAAGCCATAAACGACACACTCA  802

203 isAlaGlnIleSerArgLeuGluSer.........................  211
    :::  :|||
803 AGGGTTTCCTCTCAACCGAAGACGCCAATACCCTTTAAGTAAAAGACTTC  852

212 ..............GlyCysThrAlaThrValArgProGly...........  220
                  |||     :
853 AGAAGGGTTTTAATAATAGCGTGCGGGACCTCTTACCACCGCGGCGTTCGT  902

221 .GlyLysLeuGluGlnLysArgTyrPheLysProGlnPheProValGlnL  237
    ||||||    |||                |||:|:
903 CGGAAAGTACTGGATAGAGAGATT......GCAGGTGTTCCCACAGAGG  946

237 ...................ValLysGlyLysGluGln  244
                       |||   :|:
947 TAATTTACGCTGGAATTCAGGTATGCGAGCGTTCCCGGTTCGGACAAG  996

245 AspLeuPheAspArgIleAlaGln..........................  252
    :|:|
997 GATATCGTTATCGGAATTTCCCAGTCAGGAGAGACCGCTGACACAAAGTT  1046

253 .ValLeuGluAspSerValGluLysHisMetArgAlaAspValThrValG  269
    ::|||:|:         |||               GGAGCCTTTACCGTGG  1087
1047 TGCCCTTCAGTCCGCAAAGAAAAG.............GGAGCCTTTACCGTGG

269 ly......SerPheLeuSerGlyGlyIleAspSerThrAlaIleAlaPro  283
    ||             |||:|:|:|:||||
1088 GACTCGTAAACGTAGTGGGAAGTCCCATAGACAGGGAGTCGGACTTTTCC  1137

284 LeuAlaLysArgHisAsnProAspLeu..............LeuThrPh  295
    |||                                       :|:
1138 CTTCACACACATGCGGGACCCGAAATAGGCGTGGCGCGCTACAAAGACCTT  1187

295 eThrThrGlyPhe..................................G  300
    |||:|:|:
1188 CACCGCCACAGTTCACCGCACTCCACGCCCCTTCGGTAAGGGAAAGTGAGG  1237

300 luArgGluGly..........TyrSerGluValAspValAlaAlaGlu  312
    ||||||||          :::|||
1238 AGAGGGAAAATCTAATAAGACTCCTTGAAAAGGTTCCATCACTCGTTGAA  1287

313 SerAlaAlaAlaIleGlyIleAlaGluHisIleValLysIleValSerProGl  329
    |||:|:        |||:|:|||||
1288 CAAACACTGAACACC...GCAAGAAGTGGAGAAGGTAGCGGAAAAGTA  1334

329 uGluTyrAlaAsnAlaIleProLysIleMetTrpTyrLeuAspAspProV  346
    |||                |||:  :|||:
1335 CATGAAAAAGAAAAAACATGCTTTACCTCGGAAGGTACTTAAATTACCCCA  1384

346 alAla..........AspProSerLeuValProLeuTyrPheValAlaAla  359
    :|||                               |||
1385 TAGGCCGTGGAGGAGCTCTTAAACTTAAAGAAATTTCTTACATACACGCG  1434

360 GluAla..........ArgLysHisValLysValValLeuSerGl  371
    |||||                                :|||
1435 GAAGGTTATCCCGCAGGGAGATGAAGCACGTCCCCATAGCCCTCATAGA  1484
```

```
                                                    Quality:      154.50          Length:       580
                                                    Ratio:         0.592          Gaps:          27
                                Percent Similarity:  45.000    Percent Identity:  20.000

                                alignment_block:
                                US-09-786-474-2 x US-09-412-184-22

371 yGluGlyAlaAspGluPheGlyGlyTyrThrIleTyrLysGluProL 388       Align seg 1/1 to: US-09-412-184-22  from:  1  to: 1779
        ||:::                  ::
1485 CGAAAACATGCCGGTTGTGGTA.......................... 1506       1 MetCysGlyLeuLeuGlyIleLeuThrAlaAsnGlyAsnAlaGluAlaPh 17
                                                                     ||:||||||:::||||     ||||:::|||
388 euSerLeuAlaPro..........PheGluLysIleProSerProLeu 400       1 ATGTGCGGATTAGTCGGATACGTA..........GGGAGGGATTTAGC 38
        :::||||||                :::||||||
1507 ......ATCGCACCGAAAGACACAGGGTTTACGAGAAGATACTCAAACGTA 1551      17 eValProAlaLeuArgAlaLeuProCysMetArgHisArgGlyProA 34
                                                                     ::||||   :::||||||   ::||||:::||||||  :
401 ArgLysGlyLeuGlyLysLysLeuSerLysValLeuProArgGlyMetLysGl 417   39 CCTTCCTATAGTCCTCGGAGCTCTTGAGAGACTCGAATACAGGGGTTACG 88
        :::||||  ||:::||||||     ||:::||||
1552 GAAGAGGTTCTCCGACGAAAGGAAGGGTTATTCGTGAGGCTTTAAAGG 1601      34 spAspAlaGlyThr.............................. 38
                                                                     ||:::||||||
417 yLysSerLeuGluArgGlySerMetThrMetGluGlu 430                   89 ACTCCGCGGGAGTTGCCCTTATAGAAGACGGGAAACTCATAGTTGAAAAG 138
        :::||||||  |||           |||
1602 AGACGAAACTCAAAAGCAAATCCGAGCGTTATGGAA 1641                  39 ............................Trp..........H1 40
                                                                                                ||
                                                                    139 AAGAGGGAAAGATAAGGGAACTCGTTAAAGCCGTATGGGAAAGGATTA 188

                                                                     40 sAspAlaAspAlaAlaPheGlyAlaPheAsnArgLeuSer.............. 52
                                                                         :::|||      |:::|||
                                                                    189 CAAGGCTAAAACGGGTATAGGTCACACACGCTGGGCAACCCACGGAAAGC 238

                                                                     53 ..IleIleAspIleAlaHisSerHisGlnProLeuArgTrpGlyProAla 68
                                                                         ::  :::||||||   |||
                                                                    239 CCACGGACGGAGAACGCCCACCCCCAC.................ACC 267

                                                                     69 AspGluProAspArgTyrAlaMetThrPheAsnGlyGluIleTyrAsnTy 85
                                                                         ||||||        |:::|||  ||||||  ||:||
                                                                    268 GACGAAAAAGGTGAGTTTGCAGTAGTTCACAACGGGGATAATAGAAAACTA 317

                                                                     85 rValGluLeuArgLysLeuSerAspLeuGlyTyrThrPheAsnThrS 102
                                                                         ||:::|||    ::||||||:::||||||
                                                                    318 CTTAGAACTAAAAAGAAGACTAAAGAAGAAGAAGTGTAAAGTTCAGGTCCG 367

                                                                    102 erGlyAspGlyGluProIle......ValValGlyPheHisHisTrpGly 116
                                                                         ::  |||     |||         :::||||||
                                                                    368 AAACAGACACAGAAGTTATAGCCCACCTCATAGCGAAGAACTACAGGGGG 417

                                                                    117 Glu..........SerValValGluHisLeuArgGlyMetPh 127
                                                                         :::          ||||||      ||:::|||  ||
                                                                    418 GACTTACTGGAGGCCGTTTTAAAAACCGTAAAGAAATTAAAGGGTGCTTT 467

                                                                    127 eGlyIleAlaIleTrpAspThrLysGluLysSerLeuPheLeuAlaArgA 144
                                                                         |:::|||:::||||||      |||  :::||||||
                                                                    468 TGCCTTTGCGGTTATAACCGTTCACGAACCAAACAGACTAATA........ 510

                                                                    144 spGlnPheGlyIleLys..........ProLeuPheTyrAlaThrThrGlu 157
                                                                         |||:::|||          |||  :::
                                                                    511 ........GGAGTGAAGCAGGGAGTCCTTTAATCGTCGGACTCGGAGAA 552

                                                                    158 HisGlyThrValPheSerGluLysLysThrIleLeuGluMetAlaGl 174
                                                                         :::||||||     ::||||||
                                                                    553 GGAGAAAACTTCCTCGCTTCAGATATTCCCGAATACTTCCTTACACGAA 602

                                                                    174 uGluMet.................................... 176
                                                                         :::||||||
                                                                    603 AAAGATTATTGTTCTTGATGACGGGAAATAGCGACCTCGACTCCCGACA 652

                                                                    177 ..............AsnLeuAsp.................... 179
                                                                                    |||
                                                                    653 CTGTGAACATTTACAACTTTGAGGGAGGAGAGCCCGGTTTCAAAGGAAGTAATG 702

                                                                    180 .............LeuGlyLeuAspLysArgThrIleGluHisTy 190
                                                                                  ::: ||||||  :::||||||
                                                                    703 ATTACGCCCTGGGATCTTTGTTTCTCGCGAAAAGGGGTGGTTTAAACACTT 752
```

```
190  rValAspLeuGlnTyrValProGluPro.........AspThrLeuH  203
         :::      |||      :::|||
753  CATGCTAAAAGAGATATACGAACAGCCCAAAGCCCATAAACGACCACTCA  802

203  isAlaGlnIleSerArgLeuGluSer................      211
         :::::|||
803  AGGGTTCCTCTCAACCGAGACGCAATACCCTTTAAGTAAAGACTTC      852

212  .........GlyCysThrAlaThrValArgProGly.......      220
              :::|||      :::  |||
853  AGAAGGGTTTAATAATAGCGTGCGGACACCTCTTACCACGCCGGCCTTCGT  902

221  .GlyLysLeuGluGlnLysArgTyrPheLysProGlnPheProValGlnL  237
         |||||      :::::::::      |||::::::
903  CGGAAAGTTACTGGATAGAGAGATTT......GCAGGTGTTCCCACAGAGG  946

237  ysVal.........................ValLysGlyLysGluGln   244
         |||                              ::::|||
947  TRATTTACGCTTCGGAATTCAGGTATGCGGACGTTCCCGTTTCGGACACAAG  996

245  AspLeuPheAspArgIleAlaGln......................      252
         |||::::::|||
997  GATATCCGTYATCCGGAATTTCCCAGTCAGGAGAGACCGCTGACACACAAGTT  1046

253  .ValLeuGluSerValGluLysHisMetArgAlaAspValThrValG   269
         ::::::|||::::      |||      :::      |||||||
1047 TGCCCTTCCAGTCCGCAAAGGAAAAG......GGAGCCTTTACCGTGG  1087

269  ly.......SerPheLeuSerGlyGlyIleAspSerThrAlaIleAlaPro  283
         :::      ::::::::|||::|||||||
1088 GACTCGTAAACGTAGTGGGAAGTGCCATAGACAGGGAGTCGGACTTTTCC  1137

284  LeuAlaLysArgHisAsnProAspLeu..........LeuThrPh     295
         |||      ::::||::||||:      |||||
1138 CTTCACACACATGCGGGACCCCGAAATAGGCGTGGCGGCTACAAAGACCTT  1187

295  eThrThrGlyPhe........................G         300
         |||:::|||      ||
1188 CACCGCACAGTTCACCGGACTTCACGCCCTTTCGGTAAGGAAAGTGAGG    1237

300  luArgGluGly.........TyrSerGluValAspValAlaAlaGlu    312
         |||||||||      :::|||
1238 AGAGGGAAAATCTAATAAGACTCCTTGAAAAGGTTCCATCACTCGTTGAA  1287

313  SerAlaAlaIleGlyAlaGluHisSerIleValLysIleValSerProGl  329
         ::::::      |||||||:::::|||
1288 CAAACACTGAACACC...GCAGAAGAAGTGGAGAAGTGACCGGAAAGTA  1334

329  uGluTyrAlaAsnAlaIleProLysIleMetTrpTyrLeuAspSerProV  346
         :::      ::|||:::
1335 CATGAAAAAGAAAAACATGCTTTACCTCGGAAGTACTTAAATTACCCCA  1384

346  alAla.......AspProSerLeuValProLeuTyrPheValAlaAla    359
         :::|||      |||:::
1385 TAGGCGCTGGAGGGAGCTCTTAAACTTAAAGAAATTTCTTACATACACGCG  1434

360  GluAla.........ArgLysHisValLysValValLeuSerGl     371
         |||::::      ::::::|||
1435 GAAGGTTATCCCGCAGGGAGGATGAAGCACGTCCCATAGCCCTCATAGA   1484

371  yGluGlyAlaAspGluLeuPheGlyGlyTyrThrIleTyrLysGluProL  388
         :::      |||:::||::|||||||:::||
1485 CGAAAACATGCCCGTTGTGGTA.....                         1506

388  euSerLeuAlaPro..........PheGluLysIleProSerProLeu   400
         ::::      ::|||:::|||      |||
1507 ....ATCGCACCGGAAGACAGGGTTTACGACGAAGATACTCTCAAACGTA  1551

401  ArgLysGlyLeuGlyLysLeuSerLysValLeuProAspGlyMetLysGl  417
         :::::      ||||||      :::::|||::
1552 GAAGAGGTTCTCGCAAGAAAGGAAGGTTATTTCTGTAGGCTTTAAAGG    1601

417  yLysSerLeuLeuGluArgGlySerMetThrMetGluGlu  430
         |||||||
```

```
1602  AGACAGTAAAGAACTCTCAAAAGCAAATCCGAGAGCCGTTATGGAA  1641
```

seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:us-09-328-111-544

seq_documentation_block:
; Sequence 544, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 544
;   LENGTH: 570
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(570)
;   OTHER INFORMATION: n = A,T,C or G
US-09-328-111-544

alignment_scores:
         Quality: 152.00   Length: 198
         Ratio: 1.434   Gaps: 7
Percent Similarity: 53.535   Percent Identity: 26.263

alignment_block:
US-09-786-474-2 x US-09-328-111-544/rev

Align seg 1/1 to reverse of: US-09-328-111-544 from: 1   to: 570

```
105  GlyGluProIleValValGlyPheHisHisTrpGlyGluSerValValGl  121
         ||||      |||      :::||
545  GGTGAGATAATCCTT..........CATCTTTATGCCAAGGAGGAATTGA  505

121  uHis.........LeuArgGlyMetPheGlyIleAlaIleTrpAspT   134
         ||||      |||||      ::::|||:::
504  GCAANCAATTTGTATGTTGGATGGTGTGTTNGCATTTGTTTTACTGGATA  455

134  hrLysGluLysSerLeuPheLeuAlaArgAspGlnPheGlyIleLeuPro  150
         :::||||      :::::|||||      ::|||
454  CTGCCAATAAGAAAGTGTTCCTGGTAGAGATACATATGGAGTCAGACCT  405

151  LeuPheTyrAlaThrThrGluHisGly...ThrValPheSerSerGluLy  166
         |||      |||:::      :::::
404  TTGTTTAAAGCAATGACAGGAGATGGATTTTGGCGTGATGTTCAGAAGC  355

166  sLysThrIleLeuGluMetAlaGluGluMetAsnLeuAspLeuGlyLeuA  183
         |||      |||:::      :::
354  TAAAGGTCTTGTTACATTGAAGCACTCCGCGACTCCCTTTTTAAAAGTGG  305

183  spLysArgThrIleGluHisTyrValAspLeuGlnTyrValPro.....  197
         :::      |||
```

```
  1 MetCysGlyLeuLeuGlyIleLeuThrLeuAlaAsnGlyAsnAlaGluAlaPh  17
    |||||||||::   :   :::::            :::|||
128 ATGTGTGGCATTTGGGCGCTGTTTGGCAGTGATTGATTGCCTTTCTGCT..   175

 17 eValProAlaLeuGluArgAlaLeuProCysMetArg.....HisArgG   32
                    :::|||         ::|::::|           |||||||     |
176 ..........CAGTGTCGAGTGCTATGAAGATTGCACACAGAG        209

 32 lyProAspAspAlaGlyThrTrpHisAsp.........AlaAspAla   44
    |||| ||||||   :::|||::                           |||||
210 GTCCA...GATGCATTCCGTTTGAGAATGTCAATGGATACACCAACTGC  256

 45 AlaPheGlyPheAsnArgLeuSerIleIleAspIleAlaHisSerHisGl  61
    ::||||||||:||::|||:::||:|||::|             ::   ::  |
257 TGCTTTGGATTTCACCGGTGGCGGTAGTTGACCCGCTGTTTGGAATGCA   306

 61 nProLeuArg................TrpGlyProAlaAspGluP  71
    ||||::||                                              |||
307 GCCAATTCGAGTGAAGAAATATCCGTATTTGTG.............       340

 71 roAspArgTyrAlaMetThrPheAsnGlyGluIleTyrAsnTyrValGlu  87
         :::||||||||||||||||  :::
341 ......CTCGTTACCAATGGTGAAATCTACAACCATAAGAGAG          376

 88 LeuArgLysGluLeuSerAspLeuGlyTyrThrPheAsnThrSerGlyAs 104
    ::||||||||||||::::||::|||                   ||||||||||||||
377 ATGCAACAGCATTTGAA.........TTTGAATACCAGCACAAAGTGGA   417

104 pGlyGluProIleValGlyPheHisHisTrpGlyGluSerValValG 121
    |||||||    ||:::|:|      ::::::             :::::
418 TGGTGAG...ATAATCCTTCATCTTTATGACCAANGGAGGAATTGAGCCA  464

121 luHisLeuArgGlyMetPheGly...IleAlaIleTrpAspThr......  134
    ::||||||   |||  ::||||     ::|||||||       |||
465 ACCATTTGNATGGTTGGATGGGTGTGGTTGCAATTNGGTTTACTGGGGAA  514

135 .........LysGluLysSerPheLeuAlaArgAspGlnPheGlyIl 148
                     :::|||::::|||||||               :::|||
515 ACTGGCCATTANGAAAAGGGNTCCTGGGTAAAAGAATCCCTATGGGGGCC  564

148 eLysProLeu 151
    ||||||
565 NNAACCTTTG 574
```

seq_name: /cgn2_6/ptodata/1/1na/6B_COMB.seq:US-09-328-111-395

```
304 AGCCTTTCTTCCTCGGACACTATGAAGTTTGGATTTAAAGCCAAATGGC 255

198 ..................................GluProAspTh 201
                                                  :::   :::
254 AAAGTTGCATCCGTGGAAATGGTTAAATATCATCACTGTCGGGGATGAACC 205

201 rLeuHisAlaGlnIleSerArgLeuGluSerGlyCysThrAlaThrValA 218
    |||||||        :::::::||::|                             :::
204 CCTGCACGCCCCTCTATGACAATGTGGAGAAA..............CTCT 170

218 rgProGlyGlyLysLeuGlnLysArgTyrPheLysProGlnPhePro 234
    :::   ||||||||||||:::||
169 TTCCAGGTTTTGAGATAGAA.............................  150

235 ValGlnLysValValLysGlyLeuLysGluGlnAspPheAspArgIleAl 251
    ::||||||:||::::::::             ::   :::|||||:::
149 ........ACTGTGAAGAACAACCTCAGGATCCTTTTAAT.......... 117

251 aGlnValLeuAspSerValGluLysHisMetArgAlaAspValThrV 268
        ::::::|||::::|||  :::|||
116 ..........AATGCTGTAAAGAAACGTTTGATGACAGACAGAAGGA   80

268 aIGlySerPheLeuSerGlyGlyIleIleAspSerThrAlaIIeAla 282
    :::     :::|||||||||||:::   :::   :::|||
 79 TTGGCTGCCTTTATCAGGGGCGTTGGACTCCAGCTTGGTTGCT   36
```

```
alignment_scores:
      Quality:   132.00          Length:          170
        Ratio:    1.333            Gaps:           10
Percent Similarity:  58.235     Percent Identity:  26.471

alignment_block:
US-09-786-474-2 x US-09-328-111-395

Align seg 1/1  to: US-09-328-111-395     from: 1  to: 629
```